Predictive Processing Interpretation of the Mirror Test and Implications of a Reflection Prediction for Human Cognition

Sean O'Connor[1], Daniel Ari Friedman[2]

1) **Spoconno@gmail.com**
   https://orcid.org/0000-0002-0858-5635
   Independent Researcher
2) **DanielAriFriedman@gmail.com**
   https://orcid.org/0000-0001-6232-9096
   Active Inference Institute
   University of California Davis, Department of Entomology & Nematology

Version 1, uploaded November 29, 2022

Abstract

The "mirror test" has been used as a behavioral measure of mirror self-recognition for a variety of species. In this article we apply a predictive processing interpretation to the results of the mirror test in order to offer a novel perspective with which to understand mirror self-recognition and self-directed behavior. Furthermore, we hypothesize that a "reflection prediction", upon which our predictive processing interpretation of the mirror test is built, may also offer a novel perspective to understand how humans locate themselves relative to a mirror, imitate others, and are self-aware from a social perspective. As we show that a reflection prediction may help to explain how these traits may emerge in human cognition, we also point out that atypical reflection predictions or atypical use of a reflection prediction may help to explain instances where these traits are atypical in certain individuals.

Introduction

In “The free-energy self: A predictive coding account of self-recognition” Matthew Apps and Manos Tsakiris explain how mirror self-recognition may be understood as the result of an individual successfully learning to predict their own reflection in the mirror [1]. As they write:

> “the viewing of an agent’s own actions in a mirror (including arm movements, facial expressions etc.) will lead to optimised high level empirical priors about one’s body, which will in turn modulate expectations in the visual system about the expected visual consequences of one’s own actions.”

These optimized high level empirical priors are an individual’s expectations about their own visual reflection and therefore what we refer to here as a “reflection prediction”. In this paper we highlight that a successful reflection prediction explains mirror self-recognition in humans, as pointed out by Apps and Tsakairis, and also offers a novel perspective on how to interpret the results of the “mirror test” in other species.

Furthermore, we point out that the concept of a reflection prediction is crucial for the scientific and medical community to consider because if a reflection prediction allows agency over a reflection then reflection prediction errors may cause atypical agency over a reflection, and reflection prediction error correction may help to reverse atypical agency over a reflection.

We conclude by pointing out that a reflection prediction should also be considered as a mechanism the human brain may depend upon for the emergence and the successful functioning of cognitive traits other than mirror self-recognition such as the ability to locate oneself relative to a mirror, the ability to imitate others, and the ability to be self aware from a social perspective. As with mirror self-recognition, we point out that if a successful reflection prediction contributes to the proper functioning of these traits, then reflection prediction errors or errors with the use of a reflection prediction may cause atypical functioning of these traits as seen in certain individuals.

Predictive Processing Perspective of the Mirror Test

In 1970 Gordon Gallup Jr. devised a test to determine whether animals recognize themselves in the mirror and therefore whether they were capable of mirror self-recognition [2]. To accomplish this he anesthetized chimpanzees and monkeys, applied an odorless red dye to the animals eye ridge and opposite ear, and after a period of time observed their behavior both with and without a mirror. When the mirror was added to the experiment, chimpanzees with prior mirror experience exhibited self-directed behavior towards the mark and spent more time staring into the mirror, but monkeys with prior mirror experience did not exhibit self-directed behavior towards the mark or spent more time staring into the mirror; therefore, Gallup determined that chimpanzees were capable of mirror self-recognition and monkeys were not capable of mirror self-recognition. This test has come to be known most popularly as the “mirror test” (also, “mark

test") and it has been repeated many times across a range of species in the years since Gallup's original paper.

In this paper we will show that applying the predictive processing framework to the results of the mirror test offers a novel perspective to interpret the results of the mirror test. Predictive processing models suggest that individuals build internal predictive models of the external world and seek to minimize the extent to which their predictions differ from reality. In predictive processing models if an individual's prediction is different from their experience they can engage in prediction error minimization by either updating their prediction, a process called perceptual inference, or through acting on their environment, a process called active inference [3]. If an individual's prediction is equal to their experience, the individual's attention to the sensory experience is dampened in a process called sensory attenuation and, for self-generated movement, agency is achieved over the movement.

Applying this predictive processing framework to the results of the mirror test offers a new perspective to understand mirror self-recognition[1] and self-directed behavior, where mirror self-recognition is due to sensory attenuation when an individual successfully predicts their reflection and self-directed behavior is active inference to correct reflection prediction errors. In the next two sections we will discuss the implications of each of these claims.

Mirror Self-Recognition is a Product of Sensory Attenuation

As mentioned in the introduction of this paper, Apps and Tsakiris use a predictive coding model to describe how mirror self-recognition in humans may be considered a product of sensory attenuation when an individual successfully predicts their reflection [1]. In this section, we point out that this concept is applicable not only for humans but also animal species, offering a new way to interpret the results of the mirror test. When responding to the mirror test species fall into one of three categories - those which treat their reflection as a conspecific indefinitely, those which treat their reflection as a conspecific before ignoring their reflection entirely, those which treat their reflection as a conspecific before displaying self-directed behavior. **Traditionally only those species which display self-directed behavior are considered capable of mirror self-recognition, however, this predictive processing analysis of the mirror test offers a new perspective where both species which learn to ignore their reflection in the mirror and species which display self-directed behavior in the mirror are capable of mirror self-recognition.**

That individuals which ignore their reflection are capable of mirror self-recognition may seem strange, however, it is supported by both the predictive processing model and the ecological evidence. It is supported by the predictive processing model because sensory attenuation of self-generated behavior leads to a dampening of attention towards external sensory stimuli and has been claimed to lead to a sense of agency over that behavior [4,5], and thus help generate self-other distinction [6,7]. It is supported by ecological evidence because studies have shown

[1] And self-recognition in naturally reflective surfaces such as standing pools of water or mineral deposits in caves or on the ocean floor

that even among species which are capable of self-directed behavior, individuals may not display self-directed behavior during the mirror test. For example, one study on Asian elephants found that only one of three elephants passed the mirror test [8]. The authors therefore suggest the test may be flawed because the elephants simply may not care about the mark - as they write, “A small paint mark may be trivial to them.” Another example of this is that studies have found even for Chimpanzees, which were used in Gallup’s original experiment, fewer than half of the individuals studied pass the mirror test [9] [10]. These studies highlight an inherent flaw in the scientific interpretation of the mirror test; the mirror test measures self-directed behavior, not mirror self-recognition, and many individuals or even species who fail to exhibit self-directed behavior may nevertheless be capable of mirror self-recognition, a suggestion which is consistent with the predictive processing analysis of the mirror test presented here.

Self-Directed Behavior is an Instance of Active Inference

If self-directed behavior is not an accurate measurement of mirror self-recognition, what exactly is the self-directed behavior exhibited during the mirror test an accurate measurement of?

One possibility is what we call “mirror proprioception”, the understanding of the location of self in space relative to a mirror. As Gallup writes: “self-directed and mark-directed behaviors would seem to require the ability to project, as it were, proprioceptive information and kinesthetic feedback onto the reflected visual image so as to coordinate the appropriate visually guided movements via the mirror” [2]. We agree, however, it is important to note that self-directed behavior is also not an accurate measurement of mirror proprioception in a species for the same reason as mentioned above; the subjects being studied may simply not care about the mark. Indeed, a number of species have passed “mirror use” tests that demonstrate their ability to use a mirror to locate objects in space but have nevertheless failed to demonstrate self-directed behavior during a mirror test [11], supporting the idea species which do not display self-directed behavior may nevertheless have a visuospatial understanding of mirrors.

Therefore, self-directed behavior is an accurate measurement of neither mirror self-recognition nor mirror proprioception. In this model, self-directed behavior is conceptualized as a specific instance of active inference. Earlier we wrote that in predictive processing models, when prediction of self-generated sensory input is different from the experience of self-generated sensory input, prediction error correction has been proposed to occur either in the form of perceptual inference (updating predictions) or active inference (acting on the environment). **Applying this predictive processing framework to the results of the mirror test offers a new perspective to understand self-directed behavior as active inference to correct reflection prediction errors.** In the case of the mirror test, the mark causes a reflection prediction error because the individual's experience of their reflection now differs from their prediction of their reflection, and self-directed behavior is active inference to correct the reflection prediction error caused by the mark. This view of self-directed behavior as active inference also provides an explanation of why some individuals or even species seem to “not care” about the mark during a mirror test; those individuals or species may be ignoring that reflection prediction error because they are using perceptual inference to update their reflection

prediction in order to correct the reflection prediction error caused by the mark during the mirror test. This prediction error correction from perceptual inference would theoretically re-allow sensory attenuation because the updated prediction would allow reflection prediction to equal reflection experience and make self-directed behavior unnecessary, and atypical perceptual inference may thus explain certain instances of increased self-directed behavior.

While interpreting self-directed behavior during the mirror test as active inference is novel, the idea is consistent with the concept of active inference in predictive processing models, if we consider that the individual emerges to themselves as a part of the environment through a mirror. For example, Constant et al. argue that the self can extend to objects in the environment via niche construction, a process in active inference where an individual modifies the environment based on their internal model. As they write:

> "An agent not only learns a model of its environment, but also changes the environment to fit its models. As a result, it tends to construct an environment that mirrors its predictions, and, in a sense, makes the world its own." [12]

Thus, if an individual has emerged to themselves as a part of the environment through a mirror, the individual may exhibit self directed behavior as a way to "construct an environment that mirrors [their] predictions".

Therefore in this predictive processing interpretation of the mirror test self-directed behavior in a species is a measurement of whether the species faced a strong enough selection pressure to drive the natural selection of active inference in front of a reflective surface. This illuminates why some species capable of mirror self-recognition display self-directed behavior while other species capable of mirror self-recognition do not display self-directed behavior. Examples of species that may have faced a strong enough selection pressure to drive the natural selection of active inference in front of a reflective surface include those within which individuals are threatened directly by diseases carried by blood or parasites which may be removed with self-directed behavior, and/or species in which individuals are threatened indirectly by blood or parasites because physical appearance has a reproductive fitness consequence. For example, one of the species which has passed the mirror test by exhibiting self-directed behavior is the cleaner wrasse [13]. According to this theory, the reason the cleaner wrasse exhibits self-directed behavior is that they are threatened directly by blood or parasites and/or threatened indirectly by blood or parasites causing a reproductive fitness consequence. Considering the cleaner wrasse survives by eating parasites off of bigger fish in a mutualistic relationship, it is indeed likely that they are threatened both directly and indirectly by parasites which can be removed in front of a reflective surface; directly because they spend their lives eating them off bigger fish, and indirectly because the big fish may be less likely to enter a mutualistic relationship with cleaner wrasse that is already infected by parasites.

Reflection Prediction Errors and Error Correction

If a successful reflection prediction is responsible for agency over a reflection in the mirror as first suggested by Apps and Tsakairis, it is important for the scientific and medical community to

consider that reflection prediction errors may cause atypical mirror self-recognition as seen in a variety of individuals, and that reflection prediction error correction may help to treat atypical mirror self-recognition. Moreover, certain behaviors may be considered attempts at reflection prediction error correction analogous to the self-directed behavior observed during the mirror test, as discussed earlier.

To elaborate on this we will point out that although a reflection prediction is theoretical, one method to investigate this topic is to consider what errors a reflection prediction could have, what the effects of this error would be, and how the brain could correct this error, then compare these predictions with symptoms and behaviors actually observed in various individuals.

For example, one hypothetical reflection prediction error that we can consider is that an individual predicts their reflection is missing a limb that the individual is not actually missing. This circumstance would result in a reflection prediction error because the individual's reflection prediction would be different from their actual reflection. The individual may use perceptual inference to update their reflection prediction and correct the reflection prediction error, however, if perceptual inference fails, active inference would be necessary to correct the reflection prediction error. Active inference could take two forms - either amputate the limb so that reflection prediction is equal to experience, or use a virtual reality mirror that allows the individual to see their reflection without the limb.

The hypothetical reflection prediction error just described closely resembles body integrity identity disorder (BIID) - a condition characterized by individuals desiring amputation. The model presented in this paper offers an explanation for this experience if we consider that individuals may suffer from a reflection prediction error where they no longer predict the reflection of a certain limb. In this circumstance the desire of an individual to amputate may be considered active inference for reflection prediction error correction, analogous to the self-directed behavior mentioned earlier in this paper, because amputation will correct the reflection prediction error as the individual will now no longer see the reflection of the limb; therefore, reflection prediction and reflection experience will align and sensory attenuation may be achieved. Another way of achieving this alignment would be to use a virtual reality mirror that allows the individual to see their reflection without the limb that they no longer expect to see in the mirror. This is exactly what occurred in one study [14] that used a virtual reality mirror to “virtually amputate” specific limbs of two men with BIID; both men reacted to the virtual amputation positively, for example, one said it was “emotionally overwhelming to finally see himself like this”, while the other reported a reduction in his BIID symptoms in the days following the virtual amputation. Thus BIID closely resembles a disorder caused by a hypothetical reflection prediction error where an individual does not predict the reflection of a specific limb.

We therefore urge the scientific and medical community to consider other ways that reflection prediction errors could explain atypical mirror self-recognition and consider how reflection prediction error correction could help to treat these reflection prediction errors. If the brain is willing to amputate part of the body in order to correct a reflection prediction error, as may be the case with BIID, what other behaviors could be viewed as attempts at active inference to

correct reflection predictions errors, and what could be done in order to attempt to treat these reflection prediction errors? A few suggestions are listed below in Table 1.

Table 1

Hypothetical reflection prediction error	**Possible symptoms due to reflection prediction error**	**Possibly related disorder**	**Possible reflection prediction error correction to reduce symptoms**
Individual predicts reflection is different than reality	Claiming person in the mirror is someone else, compulsive plastic surgery	Mirrored-self misidentification	VR mirror or plastic surgery that allows reflection to more closely resemble reflection prediction
Individual predicts reflection weighs less than reality	Compulsively losing weight	Anorexia, bulimia	VR mirror that shows reflection that weighs less than in reality
Individual predicts reflection is bigger than reality	Compulsively gaining weight	Compulsive overeating	VR mirror that shows reflection that weighs more than in reality
Individual predicts reflection looks dead	Emulating the dead, suicidal behavior	Cotards disorder, depression	VR mirror that shows reflection looks dead
Individual predicts reflection is in pain	Pain	Psychogenic pain	VR mirror that shows reflection in pain
Individual predicts reflection looks paralyzed	Paralysis	Psychogenic paralysis	VR mirror that shows reflection paralyzed

Table 1. Hypothetical reflection prediction errors and their possible associated symptoms, disorders, and methods for error correction.

Implications of Reflection Prediction for Human Cognition

This paper has focused on the implications of a predictive processing analysis on the mirror test and the implications of a reflection prediction for mirror self-recognition and atypical mirror self-recognition. Here we shift focus to consider that the brain may use a reflection prediction for purposes other than mirror self-recognition; specifically but not exclusively, for locating oneself in space relative to a mirror, for imitation, and for self-awareness. As with mirror self-recognition, we also consider that if each of these traits depends upon a successful reflection prediction, reflection prediction errors or errors with the use of a reflection prediction may help to understand instances where these traits are atypical in certain individuals.

Reflection Predictions and Mirror Proprioception

In addition to mirror self-recognition, a reflection prediction may provide an individual with what we refer to as mirror proprioception, the ability of an individual to locate themselves in space relative to a mirror. A visuospatial relationship between self and reflection as a product of a successful reflection prediction makes sense because in order for an individual to be able to predict their reflection they have to understand how their reflection reacts to their motor movements, for example, that as they move their hand further from themselves and closer to a mirror, their reflection will move its hand further from itself and grow in size as it appears to move towards the individual from inside the mirror. Mirror proprioception is necessary for everyday tasks like shaving, putting on makeup, and driving. If the concept of mirror proprioception does not seem intuitive, consider that species such as bats and dolphins use echoes to locate themselves in space and that echos are a reflection of sound just as mirror reflections are a reflection of light. Thus the suggestion that individuals use mirror reflections to locate themselves in space is situationally analogous to bats and dolphins using echoes to locate themselves in space.

One experiment that may demonstrate mirror proprioception is the “mirror illusion” [15]. During the mirror illusion a mirror is placed on an individual’s midsagittal axis and the individual synchronically moves their left and right hands, but cannot see one hand because of the mirror. For example, if the mirror is placed such that it reflects the left side of the individual’s body, the individual will not be able to see their right hand; instead, they see the left hand and the reflection of their left hand. According to this study, as long as the actual hand and mirror illusion hand are within 10 cm of one another, a phenomenon known as proprioceptive drift occurs where the visual information from the illusory hand in the mirror overrides normal proprioception and the individual’s perception of their body-part position changes.

For an individual to feel ownership and agency over the hand in the mirror may be expected because our model claims an individual feels agency over their reflection when reflection prediction is the same as reflection experience. In the mirror illusion, the reflection of an individual’s left hand with a mirror on their midsagittal axis gives the individual the illusion that their left hand is their right hand in the mirror. Moreover, in this experiment the individual was synchronously moving their right hand with their left hand. Therefore, in our model the individual would be predicting the movement of their right hand’s reflection and actually seeing a reflection that looks what their right hand’s reflection would actually look like (the reflection of their left hand) which would allow reflection prediction to equal reflection experience, granting the individual ownership over the reflection where they feel like it is the reflection of their right hand and agency over the reflection because they now predict the reflection will respond to the motor commands of their right hand. As one subject in this experiment said: ”Whenever I moved my right hand, I expected the hand in the mirror to move in the same way.”

Additionally, proprioceptive drift is expected because we claim a successful reflection prediction may allow mirror proprioception, the ability of an individual to locate themselves in space relative to a mirror. Therefore, during the mirror illusion when an individual feels agency over the illusory hand in the mirror, we would expect the individual may feel confused about their location

in space relative to the mirror as proprioception and mirror proprioception collide. This is consistent with the experience of proprioceptive drift as long as the mirror illusion hand and actual hand are within a certain physical distance of one another - as the actual hand moves further from the mirror illusion hand, experiences of proprioceptive drift and changes in ownership/agency are less likely to occur; in this model, this may occur because distances makes it tougher to convince the brain that the movements of the mirror illusion hand are due to the right hand.

If a successful reflection prediction indeed facilitates mirror proprioception, reflection prediction errors or errors with the use of a reflection prediction may cause atypical mirror proprioception. For example, individuals may not be able to perform everyday visuospatial tasks in front of a mirror, or may be mirror aversive and avoid environments with mirrors entirely. One example of this may be seen in a mirror drawing task where individuals are asked to trace shapes while visually focusing on their reflection in a mirror. One study has shown that some individuals with autism spectrum disorder (ASD) perform better at this task than neurotypical individuals but others perform much worse [16].The increased variance in the performance of this task in ASD individuals may reflect the heterogeneity in the underlying mechanisms that lead to a common diagnosis of ASD, however, according to our model, performing worse would be expected if some individuals had atypical mirror proprioception and struggled to successfully locate themselves in space relative to the mirror.

Additionally, atypical mirror proprioception due to reflection prediction errors or errors with the use of a reflection prediction may cause an individual to feel located in space somewhere other than where they are really located. For example, if an individual predicts their reflection is half the size it actually is this model suggests the individual be confused about where they are actually located in space relative to the mirror, and that such a reflection prediction error could be treated with a VR mirror that shows the individual's reflection as half the normal size. If reflection prediction errors or errors with the use of a reflection prediction do cause an individual to predict that they are somewhere in space other than where they are actually located in this way, this could help to explain out-of-body experiences as seen in a variety of individuals. As one patient with depersonalization disorder explains:

> “I look in the mirror and it doesn’t feel like myself I’m looking at. It’s like I’m floating, not actually experiencing the world, and slowly fading away into nothing. It’s like I’m on autopilot in somebody’s else body [17](see also [18][19]).” [6]

Therefore, we argue a reflection prediction may allow an individual to understand their location in space relative to a mirror, and failures in a reflection prediction may cause an individual to not know where they are located or feel located somewhere other than where they really are located.

In this section we have discussed how a reflection prediction error may contribute to the scientific understanding of mirror self-proprioception. We now turn to discuss how a reflection prediction may contribute to the scientific understanding of imitation.

Reflection Predictions and Imitation

The ability of a species to imitate others is also a trait that may depend on the ability of individuals within that species to successfully predict their own reflection. Unlike mirror self-recognition and mirror self-proprioception, imitation is not proposed as an automatic product of a successful reflection prediction, but rather the product of natural selection shaping the brain to use the reflection prediction for imitation. The ability of an individual to predict their own reflection before or without seeing it suggests that the reflection prediction is an association between the motor neurons for an action and the sight of an individual's reflection making the contralateral movement. **As an association between the motor neurons for an action and the sight of another (the reflection) performing the contralateral action, an individual may use the same association in reverse in an inversion of the predictive model** [20] **to generate the motor neurons for an action given the sight of another performing that contralateral action.** In other words, a reflection prediction may facilitate the imitative function of mirror neurons, defined as neurons that react both when an action is performed and when it is observed.

An objection to this hypothesis that we anticipate is that children imitate others before they recognize themselves in the mirror and therefore imitation cannot depend on a reflection association. However the association between the motor neurons for an action and the sight of another acting may allow for children to substitute others for their reflection before they are capable of recognizing themselves in the mirror. Therefore while the evolution of imitation may depend on the emergence of a reflection prediction, imitation itself does not necessarily depend on the emergence of a reflection prediction developmentally, rather, individuals may also develop this reflection prediction through observing and interacting with others. One example supporting this claim is that "being imitated" can improve imitation skills [21,22] in toddlers. A second example of this interpersonal association is that infants attempt to wipe something off their faces when watching others wipe something off their faces [23,24]. A third example of this interpersonal association is the enfacement illusion, where an individual is subjected to a tactile sensation while witnessing someone else experience the sensation. This synchronic motor experience simulates a mirror experience and causes individuals to respond to self-other distinction tasks as though the other's face was more like their own face [1,25–27], and causes individuals to report that the others face was experienced as more like "me" than before they were before the enfacement illusion [28]. According to our model this may be expected because the synchronic motor experience may cause the individual to process the other as though the other were their own reflection. This is also one possible alternate explanation for the rubber hand illusion, where an individual watches a rubber hand undergo the same tactile experience as their hidden hand and begins to feel agency over the rubber hand.

The idea that children may begin to establish their reflection association via others before mirror self-recognition suggests additionally that children may build a prediction of themselves through those they imitate. This can help to explain how individuals who have never had mirror exposure may still have a prediction of themselves to use for imitation and other traits which may depend upon a reflection prediction.

If imitation depends upon a successful reflection prediction, errors with a reflection prediction or with the use of a reflection prediction may cause atypical imitation. For example, one hypothetical reflection prediction error is a non-reversed reflection prediction. A non-reversed reflection prediction would cause an individual to predict that as they waved their right arm their reflection would wave its right arm rather than its left arm, as is actually the case in front of a mirror.

If an individual has associated the movements of the left side of their body with the image of the left side of the body of others, rather than the right side of the body of others as with a normal reflection, then the individual may imitate the other incorrectly - if they attempt a contralateral imitation, they will perform an ipsilateral imitation and vice versa. One disorder which exhibits symptoms resembling this is autism spectrum disorder. Studies have indicated that autistic individuals have deficiencies with mirror-image imitation [29], and according to one first person account this sort of reversed imitation is consistent with the experience of some autistic individuals. As Willey [30] writes,

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

Therefore, while we are not suggesting a non-reversed reflection prediction is the sole cause of ASD, these specific symptoms do closely match those that would be expected if an individual did have a non-reversed reflection prediction that caused backwards imitation. One way to attempt prediction error correction would be through the use of a backwards "anti-mirror". An "anti-mirror" is a full-length computer screen that can display one's reflection as a mirror would. A backwards "anti-mirror" will cause the individual's reflection to perform the ipsilateral/anatomical rather than contralateral/mirror movement. If our claim that some individuals with autism have a backwards reflection prediction is correct then a backwards "anti-mirror" should allow these individuals to feel more agency (evaluated using self-report or other accepted neurobehavioral measures) than a normal mirror because the backwards anti-mirror allows non-reversed reflection prediction to equal reflection experience, granting the individual agency over their reflection via sensory attenuation. If true it is unclear whether this prediction error correction would reverse backwards imitation in those who exhibit this trait, however, there should be a clear positive correlation between those who exhibit backwards imitation and those who feel more agency in front of a backwards anti-mirror.

In this section we have discussed how a reflection prediction may contribute to our understanding of imitation. We now turn to consider how a reflection prediction may contribute to our understanding of self-awareness.

Reflection Predictions and Self-Awareness

One final trait we will argue may depend upon a reflection prediction is self-awareness. Like imitation but unlike mirror self-recognition and mirror self-proprioception, self-awareness is not proposed as an automatic product of a successful reflection prediction, but rather the product of natural selection shaping the brain to use the reflection prediction for self-awareness.

Many have implied a relationship between mirror self-recognition and self-awareness, suggesting that If an individual is capable of self-directed behavior that they are self-aware. Yet this claim does not stand up to logical scrutiny. For example, ants and cleaner wrasse fish exhibit self-directed behavior in front of a mirror [31][13]. Are they self aware? Bats don't exhibit self-directed behavior but they do recognize their own echo [32], which is a reflection of sound akin to how a mirror offers a reflection of light. Are bats self aware? Dog's recognize their own scent; does this mean they are self aware [33]? These rhetorical questions are of course impossible to answer without defining self-awareness. Gallup et al. write:

> "In its most rudimentary form self-awareness is the ability to become the object of your own attention. When you see yourself in a mirror, you are literally the object of your own attention, but most organisms respond to themselves in mirrors as if confronted by another organism. The ability to correctly infer the identity of the image in the mirror requires a pre-existing sense of self on the part of the organism making that inference. Without a sense of self, how would you know who you were seeing when confronted with your reflection in a mirror?" [34]

Yet as we showed above, in a predictive processing model an individual simply needs to be able to predict their reflection in order to feel agency over their reflection. This implies self-recognition, mirror and otherwise, does not require a sense of self as suggested by Gallup et al., and indeed rather suggests self-recognition may be an evolutionary precursor for a sense of self.

To explain this we suggest adding one word to Gallup's definition of self-awareness: self-awareness is the ability to become the object of your own **social** attention. This correlates with the psychological concept of objective self-awareness, defined as an attention focused on the self as a social object [35], and also in our opinion is more consistent with the colloquial use of the term self-awareness than Gallup's definition. A reflection prediction offers the potential for an individual to become the object of their own social attention because their own physical behavior has emerged to themselves in a third person perspective; therefore, they have a mechanism with which they may react to their behavior the same way they would react to the behavior of another. George Herbert Mead argued an individual becomes self aware only when they "take the attitude of the other" on their own behavior. As he writes:

> "When the response of the other becomes an essential part in the experience or conduct of the individual; when taking the attitude of the other becomes an essential part in his behavior- then the individual appears in his own experience as a self; and until this happens he does not appear as a self." [36]

Mead and others have suggested that only modes of communication which are naturally reflexive such as speech can allow an individual to take the attitude of the other because non-reflexive communication such as facial gestures cannot be processed similarly by an actor as it is by an observer [37], however, we propose that a reflection prediction can bridge this theoretical gap because evolution may allow a reflection prediction to become reflexive, allowing an individual's own gestural behavior to emerge to themselves as though they were watching themselves perform the behavior from the perspective of another, offering a mechanism for the individual to take the attitude of the other on their own behavior and therefore emerge as a social other to themselves, allowing objective self-awareness.

Combining the concept of a reflexive reflection prediction with Mead's logic suggests an individual may use their reflection prediction to react to their own behavior the same way they would react to another performing the behavior. This reaction then can be used by the individual to predict the internal reaction of others to the individual's own behavior. G.H. Mead would say the gesture has become a "significant gesture" because the gesture means the same thing to both the actor and witness. **We refer to this reaction as the "witness" reaction because the actor is proposed to react due to becoming witness to their own behavior via a reflexive reflection prediction, as an actual witness would witness their behavior.** The witness reaction implies the brain may have "witness neurons" which will appear similarly as "mirror neurons" because they are both proposed to be simultaneously activated neurons which will be shared between an actor and a witness; as mirror neurons are proposed to occur when a witness activates the same neurons as an actor to simulate the actor, witness neurons are proposed to occur when an actor predictively reacts to their behavior the same way a witness may react to the individuals behavior to simulate the witness. This has at least three important implications in regards to self-awareness.

1) Individuals may use the witness reaction for action monitoring, defined as the process for evaluating the appropriateness of one's actions [38], and response execution, the process of performing those actions, and combine these two processes as a mechanism for action simulation, simulating the reaction of others. As an individual can imitate others, individuals may perform response execution during this process of action simulation by imitating the reflection prediction.that arouses in them the reaction they wish to arouse in others.

2) When an individual chooses which behavior to perform based on the internal reaction they wish to arouse in others, this internal reaction has become the intention of the behavior. Therefore we propose during action simulation the individual is associating the intention of their behavior with the sight of another (their reflection) performing that behavior; furthermore, that individuals may use this association between intention and reflection prediction in reverse via an inversion of the predictive model [20] to infer the intention others are attempting to arouse in them.

3) The individual has emerged as another to themselves, therefore, as individuals infer the intentions of others, they may infer their own intentions during the witness reaction, as they predict others thoughts, they may predict their own thoughts during the witness

reaction, as they empathize with others, they may empathize with themselves during the witness reaction. As they narrate what others are doing, they may narrate what they themselves are doing during the witness reaction. Etc. Therefore the intention of an individual's behavior is not limited to one type of internal reaction; for example, an individual may use the witness reaction to intend to arouse thoughts in another or may intend to arouse emotions in another. Herbert Blumer, one of Mead's students, would call this a "process of indication". As he writes:

> "A very different view of human society is introduced when one recognizes that the initiator and the recipient of a significant gesture indicate the meaning of the gesture and then indicate the action to be taken on the basis of that meaning. This process of indication, allowing the possibility of choice and diverse response, breaks what would otherwise be the direct determination of action by society. It changes the position of the initiator and the recipient of a significant gesture from being mere agents of a transmitting society to the position of being actors who are confronting, defining, and handling their situations."[39]

Taken together, these implications of the witness reaction provide an explanatory mechanism to better understand not only self-awareness but also; theory of mind, defined as the ability of an individual to attribute mental states such as intentions, beliefs, thoughts, to others and themselves; mentalizing, defined as "the capacity to reflect on and interpret one's own behavior and that of others based on intentional internal mental states, such as beliefs, thoughts, and emotions" [40,41]; and thinking through other minds (TTOM), defined as "the process of inferring other agents' expectations about the world and how to behave in social context" [42]. Furthermore, these implications show that it may not be accurate to think of objective self-awareness and theory of mind as two separate concepts rather than two sides of the same coin, because after an individual emerges to themselves as an other via a reflexive reflection prediction's witness reaction, self and other are inextricably linked; as an individual understands and can define others better, they may understand and define themselves better, and as an individual understands and defines themselves better, they may understand and define others better. For example, as an individual learns to associate a name with the faces of others, they may have learned to associate their own name with their own face. As they develop feelings about another, they may develop feelings about themselves (resulting in positive or negative self-esteem). As they judge the actions of others by a moral code, they may judge their own actions by a moral code, and understand that others are judging their own actions by a moral code. Etc. Therefore, self-awareness may be considered using other-awareness on oneself, and theory of mind may be considered using self-awareness on others.

Finally, if a reflection prediction facilitates self-awareness then a reflection prediction error or the error in the use of a reflection prediction may cause atypical self-awareness. For example, if an individual has a reflection prediction that is not reflexive, that individual may not be able to predict how others will react to their behavior and may therefore demonstrate behavior that is inappropriate for a given social situation. Furthermore, since a reflexive reflection prediction is proposed as a necessary mechanism for an individual to associate their own intentions with

their reflection, an individual without a reflexive reflection prediction may not be able to use this association in reverse and therefore may struggle to understand the intentions of others.

This view of atypical self-awareness offers a new way of understanding atypical theory of mind and intention understanding as seen in certain individuals. For example, there have been multiple theories implicating mirror neuron dysfunction in intention understanding for individuals with autism spectrum disorder [43]. However, witness neurons are proposed to appear similarly in the brain as mirror neurons and witness neurons offer an easy explanation of intention understanding, as presented in this paper. Therefore, researchers may be attributing atypical intention understanding to atypical mirror neurons when they should instead be attributing atypical intention understanding to atypical witness neurons. It is therefore important to clearly differentiate between which brain processes involve mirror neurons and which involve witness neurons.

Another example of how a reflection prediction error or the error in the use of a reflection prediction may cause atypical self-awareness is if an individual consciously processes their reflection prediction as another individual this may cause them to conceptualize their predicted or actual behavior as the actual behavior of another. Two examples of individuals conceptualizing their predicted behavior as that of another may include auditory verbal hallucinations, when an individual does not recognize their thoughts and therefore claims they are the thoughts of another, and syndrome of subjective doubles, where an individual claims they have a doppelganger that is leading it's own life.

One example of an individual conceptualizing their actual behavior as that of another may be instances of multiple personalities as seen in dissociative identity disorder. An example of this can be seen in the case of Donna Williams, who has been diagnosed with autism and dissociative identity disorder and has two alternate personalities; Carol and Willie. The experience of being Carol started with atypical mirror self-recognition, as indicated by this autobiographical quote:

> "Carol came in through the mirror. Carol looked just like me, but the look in her eyes betrayed her identity. It was Carol all right. I began to talk to her, and she copied me. I was angry. I didn't expect her to do that. My expression asked her why, and hers asked me. I figured that the answer was a secret." [44]

This experience supports the suggestion that a reflection prediction error or error with the use of a reflection prediction may cause atypical self-awareness. Another example of an individual conceptualizing their actual behavior as that of another due to a mirror is seen in how a mirror can induce instances of asomatognosia in stroke recovery patients [45]. Asomatognosia occurs when an individual loses recognition of one half of their body. One common symptom of asomatognosia is somatoparaphrenia, where the individual loses agency over this part of their body. In our model this may be expected if the mirror causes a reflection prediction error on half of their reflection. In this instance the individual will not be able to successfully predict half of their reflection, so they may not be able to recognize half of their body relative to mirrors and thus may have a failure of self-awareness for that half of their body.

To summarize: a reflection prediction may facilitate self-awareness by letting an individual emerge as another to themselves, and reflection prediction errors or error with the use of a reflection prediction may cause atypical self-awareness as seen in a variety of individuals.

Conclusion

In this paper we have presented a predictive processing analysis of the mirror test, argued that a reflection prediction may facilitate mirror self-recognition, mirror proprioception, imitation, and self-awareness, and examined how reflection prediction errors or errors with the use of a reflection prediction may contribute to atypical mirror self-recognition, atypical mirror proprioception, atypical imitation, and atypical self-awareness.

Though grounded firmly in the predictive processing model built by others, this paper is largely speculative and we have offered novel perspectives on a variety of topics that have been discussed for decades or longer. These perspectives may wind up being correct or incorrect; regardless, we believe an open discussion of these perspectives can greatly advance the scientific understanding of the topics discussed in this paper such as mirror self-recognition and self-awareness. This is particularly important due to the sheer number of symptoms of disorders that this model may help to understand and possibly even treat, as we have attempted to point out throughout this paper.

Many questions remain concerning this theory. Some examples include:

- Where and how is the reflection prediction represented or implemented in the brain? Is it separate entirely from the conscious body image or are they related in some way?
- To what degree are individuals consciously aware of altered reflection predictions (i.e. would someone with BIID tell you that when they closed their eyes and pictured their reflection in the mirror, it is missing a limb)?
- What exactly is a reflection prediction error? How or in what circumstances can a reflection prediction be altered?
- What causes perceptual inference to vary or fail?
- Is there a way to treat reflection prediction errors other than reflection prediction error correction? Is there a quantitative way to measure agency over a reflection in the brain?

As a first next step we urge the scientific community to gather more data and think more critically about mirror self-recognition. The inability of an individual to recognize themselves in the mirror is currently viewed by the scientific literature as merely a symptom of various disorders and because of this there is a shortage of data on mirror self-recognition in the scientific literature. If the scientific community instead considers that an individual not recognizing themselves in the mirror may be tied to the atypical functioning of other traits, as this model suggests, we believe significant progress can be made in understanding both individual symptoms/disorders and how various symptoms/disorders relate to one another.

Finally we will conclude by pointing out that much of the model that we have presented in this paper can be considered analogously for reflexive modes of communication as a path towards

self-recognition and self-awareness under the correct evolutionary circumstances. For example, a species that communicates with odors may smell their own odor and treat their own odor as the odor of another, learn to predict how that other will respond to them, gain agency over the odor and ignore it via sensory attenuation, and use the odor prediction reflexively to generate a witness reaction in order to predict how others will respond to their own odor. A species that communicates with sound may treat their speech (or echo) as the speech of another, learn to predict how that other will respond to them, gain agency over the speech (or echo) and ignore it via sensory attenuation, and use the speech prediction (or echo prediction) reflexively to generate a witness reaction in order to predict how others will respond to their own speech. Etc. These modes of communication are different from bodily movement because they are naturally reflexive and so this seems like a likely path for self-recognition and self-awareness to arise in other species under the correct evolutionary circumstances.

References

1. Apps MAJ, Tsakiris M. The free-energy self: a predictive coding account of self-recognition. Neurosci Biobehav Rev. 2014;41: 85–97. doi:10.1016/j.neubiorev.2013.01.029

2. Gallop GG Jr. Chimpanzees: self-recognition. Science. 1970;167: 86–87. doi:10.1126/science.167.3914.86

3. Perrykkad KS. Self in Autism: A Predictive Perspective. [cited 1 Mar 2022]. Available: https://scholar.archive.org/work/yv36ywkqj5h45ieyyefzhqbvgi/access/wayback/https://au-east.erc.monash.edu.au/fpfiles/28688580/Thesis_Perrykkad_FINAL.pdf?AWSAccessKeyId=e00145a6f706457aab45051570081e49&Expires=1625774089&Signature=4OBqrrgdc%2B%2BRi%2FTORAHLPb5vpQI%3D

4. Leptourgos P, Corlett PR. Embodied Predictions, Agency, and Psychosis. Front Big Data. 2020;3: 27. doi:10.3389/fdata.2020.00027

5. Gallagher I I. Philosophical conceptions of the self: implications for cognitive science. Trends Cogn Sci. 2000;4: 14–21. doi:10.1016/s1364-6613(99)01417-5

6. Ciaunica A, Seth A, Limanowski J, Hesp C, Friston K. I overthink—therefore I am not: An Active Inference Account of Altered Sense of Self and Agency in Depersonalisation Disorder. 2021. doi:10.31234/osf.io/k9d2n

7. Haggard P. Sense of agency in the human brain. Nat Rev Neurosci. 2017;18: 196–207. doi:10.1038/nrn.2017.14

8. Plotnik JM, de Waal FBM, Reiss D. Self-recognition in an Asian elephant. Proc Natl Acad Sci U S A. 2006;103: 17053–17057. doi:10.1073/pnas.0608062103

9. Swartz KB, Evans S. Not all chimpanzees (Pan troglodytes) show self-recognition. Primates. 1991;32: 483–496. doi:10.1007/BF02381939

10. Povinelli DJ, Rulf AB, Landau KR, Bierschwale DT. Self-recognition in chimpanzees (Pan troglodytes): Distribution, ontogeny, and patterns of emergence. Journal of Comparative Psychology. 1993. pp. 347–372. doi:10.1037/0735-7036.107.4.347

11. Heschl A, Burkart J. A new mark test for mirror self-recognition in non-human primates. Primates. 2006;47: 187–198. doi:10.1007/s10329-005-0170-8

12. Constant A, Bervoets J, Hens K, Van de Cruys S. Precise Worlds for Certain Minds: An Ecological Perspective on the Relational Self in Autism. Topoi. 2020;39: 611–622. doi:10.1007/s11245-018-9546-4

13. Kohda M, Hotta T, Takeyama T, Awata S, Tanaka H, Asai J-Y, et al. Cleaner wrasse pass the mark test. What are the implications for consciousness and self-awareness testing in animals? bioRxiv. 2018. p. 397067. doi:10.1101/397067

14. Turbyne C, Koning P de, Zantvoord J, Denys D. Body integrity identity disorder using augmented reality: a symptom reduction study. BMJ Case Rep. 2021;14. doi:10.1136/bcr-2020-238554

15. Tajima D, Mizuno T, Kume Y, Yoshida T. The mirror illusion: does proprioceptive drift go hand in hand with sense of agency? Front Psychol. 2015;6: 200. doi:10.3389/fpsyg.2015.00200

16. Salowitz NMG, Eccarius P, Karst J, Carson A, Schohl K, Stevens S, et al. Brief report: visuo-spatial guidance of movement during gesture imitation and mirror drawing in children with autism spectrum disorders. J Autism Dev Disord. 2013;43: 985–995. doi:10.1007/s10803-012-1631-8

17. Perkins J. Life on Autopilot: A Guide to Living with Depersonalization Disorder. Jessica Kingsley Publishers; 2021. Available: https://play.google.com/store/books/details?id=dYofEAAAQBAJ

18. Simeon D, Abugel J. Feeling Unreal: Depersonalization Disorder and the Loss of the Self. Oxford University Press; 2006. Available: https://play.google.com/store/books/details?id=M-Cgs_T7Cl8C

19. Sierra M. Depersonalization: A New Look at a Neglected Syndrome. Cambridge University Press; 2009. Available: https://play.google.com/store/books/details?id=qOat1nDGZAkC

20. Friston KJ, Preller KH, Mathys C, Cagnan H, Heinzle J, Razi A, et al. Dynamic causal modelling revisited. Neuroimage. 2019;199: 730–744. doi:10.1016/j.neuroimage.2017.02.045

21. McDuffie A, Turner L, Stone W, Yoder P, Wolery M, Ulman T. Developmental correlates of different types of motor imitation in young children with autism spectrum disorders. J Autism Dev Disord. 2007;37: 401–412. doi:10.1007/s10803-006-0175-1

22. Ingersoll B. The Social Role of Imitation in Autism: Implications for the Treatment of Imitation Deficits. Infants Young Child. 2008;21: 107. doi:10.1097/01.IYC.0000314482.24087.14

23. Lewis M, Brooks-Gunn J. Toward a theory of social cognition: The development of self. New Dir Child Adolesc Dev. 1979;1979: 1–20. doi:10.1002/cd.23219790403

24. Johnson DB. Self-recognition in infants. Infant Behav Dev. 1983;6: 211–222. doi:10.1016/S0163-6383(83)80028-9

25. Tsakiris M. Looking for myself: current multisensory input alters self-face recognition. PLoS One. 2008;3: e4040. doi:10.1371/journal.pone.0004040

26. Sforza A, Bufalari I, Haggard P, Aglioti SM. My face in yours: Visuo-tactile facial stimulation influences sense of identity. Soc Neurosci. 2010;5: 148–162. doi:10.1080/17470910903205503

27. Maister L, Cardini F, Zamariola G, Serino A, Tsakiris M. Your place or mine: shared sensory experiences elicit a remapping of peripersonal space. Neuropsychologia. 2015;70: 455–461. doi:10.1016/j.neuropsychologia.2014.10.027

28. Tajadura-Jiménez A, Grehl S, Tsakiris M. The other in me: interpersonal multisensory stimulation changes the mental representation of the self. PLoS One. 2012;7: e40682. doi:10.1371/journal.pone.0040682

29. Avikainen S, Wohlschläger A, Liuhanen S, Hänninen R, Hari R. Impaired mirror-image imitation in Asperger and high-functioning autistic subjects. Curr Biol. 2003;13: 339–341. doi:10.1016/s0960-9822(03)00087-3

30. Willey LH. Pretending to be Normal: Living with Asperger's Syndrome (Autism Spectrum Disorder) Expanded Edition. Jessica Kingsley Publishers; 2014. Available: https://play.google.com/store/books/details?id=Q7RNBAAAQBAJ

31. Cammaerts Tricot M-C, Cammaerts R. Are ants (Hymenoptera, Formicidae) capable of self recognition ? J Sci. 2015;5: 521–532. Available: https://difusion.ulb.ac.be/vufind/Record/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/219269/Details

32. Amichai E, Blumrosen G, Yovel Y. Calling louder and longer: how bats use biosonar under severe acoustic interference from other bats. Proc Biol Sci. 2015;282: 20152064. doi:10.1098/rspb.2015.2064

33. Horowitz A. Smelling themselves: Dogs investigate their own odours longer when modified in an "olfactory mirror" test. Behav Processes. 2017;143: 17–24. doi:10.1016/j.beproc.2017.08.001

34. Gallup GG Jr, Anderson JR, Shillito DJ. The mirror test. The cognitive animal: Empirical and theoretical perspectives on animal cognition. 2002; 325–333. Available: https://courses.washington.edu/ccab/Gallup%20on%20mirror%20test.pdf

35. Duval S, Wicklund RA. A theory of objective self awareness. 1972;238. Available: https://psycnet.apa.org/fulltext/1973-26817-000.pdf

36. Mead GH, Huebner DR, Joas H. Mind, self, and society: The definitive edition. 1st ed. Morris CW, editor. Chicago, IL: University of Chicago Press; 2015. doi:10.7208/chicago/9780226112879.001.0001

37. Wiley N. Reflexivity in George Herbert Mead. Studies in Symbolic Interaction. 2021. pp. 61–72. doi:10.1108/s0163-239620210000052005

38. Suzuki K, Kita Y, Sakihara K, Hirata S, Sakuma R, Okuzumi H, et al. Uniqueness of action monitoring in children with autism spectrum disorder: Response types and temporal aspects. J Clin Exp Neuropsychol. 2017;39: 803–816. doi:10.1080/13803395.2016.1266308

39. Blumer H. George Herbert Mead and Human Conduct. Rowman Altamira; 2004. Available: https://play.google.com/store/books/details?id=u3BOxaRysbYC

40. Fonagy P, Target M, Steele H, Steele M. Reflective-functioning manual, version 5.0, for application to adult attachment interviews. London: University College London. 1998;10. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.466.3207&rep=rep1&type=pdf

41. Gergely G, Fonagy P, Jurist E. Affect regulation, mentalization, and the development of the self. International Journal of. 2002. Available: http://publications.ceu.hu/publications/gergely/2002/10159

42. Veissière SPL, Constant A, Ramstead MJD, Friston KJ, Kirmayer LJ. Thinking through

other minds: A variational approach to cognition and culture. Behav Brain Sci. 2019;43: e90. doi:10.1017/S0140525X19001213

43. Yates L, Hobson H. Continuing to look in the mirror: A review of neuroscientific evidence for the broken mirror hypothesis, EP-M model and STORM model of autism spectrum conditions. Autism. 2020;24: 1945–1959. doi:10.1177/1362361320936945

44. Williams D. Nobody Nowhere: The Remarkable Autobiography of an Autistic Girl. Jessica Kingsley Publishers; 2009. Available: https://play.google.com/store/books/details?id=MD4pDgAAQBAJ

45. Paysant J, Beis JM, Le Chapelain L, André JM. Mirror asomatognosia in right lesions stroke victims. Neuropsychologia. 2004;42: 920–925. doi:10.1016/j.neuropsychologia.2003.12.001