The active Bayesian brain and the Rorschach task

Gregory J. Meyer

University of Toledo

Karl J. Friston

The Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology

Author Note: We thank Daniel Collerton and the editor for their helpful recommendations on a previous version of this article. Correspondence concerning it may be sent to either author at gregory.meyer@utoled.edu or k.friston@ucl.ac.uk. The first author declares a potential conflict of interest in that he is a member of a company that sells the manual to the Rorschach Performance Assessment System, which is indirectly referenced in this work. The second author declares no potential conflict of interest.

Abstract

The Rorschach offers a unique and interesting paradigm from the perspective of the (Bayesian) brain. This contribution to the cross-disciplinary special issue considers the Rorschach from the perspective of perceptual inference in the brain and how it might inform subject-specific differences in perceptual synthesis. Before doing so, we provide a broad overview of active inference in its various manifestations. In brief, active inference supposes that our perceptions are the best hypothesis to explain sensory impressions. On a Bayesian account, the requisite belief updating rests sensitively upon the precision or confidence ascribed to sensory input, relative to prior beliefs about the causes of sensations. This focus – on the balance between sensory and prior precision – has been a useful construct in both cognitive science (e.g., as a formal explanation for attention) and neuropsychology (e.g., as a formal explanation for aberrant or false inference in hallucinations). In this setting, false inference is generally understood as abnormally high precision afforded to high-level hypotheses or explanations for visual input, which may compensate for a failure to attenuate sensory precision. On this view, the Rorschach offers an interesting paradigm because the amount of precise information about the causes of visual input is deliberately minimized — and rendered ambiguous — thereby placing greater emphasis on prior beliefs entertained the respondent. We close by exploring this issue and several other areas of intersection between Rorschach responding and active inference.

The active Bayesian brain and the Rorschach task

This essay starts off broadly, providing readers with the backdrop for active inference, focusing on how active inference is founded on Bayesian probabilities[1] and functions as a formal model for how human minds work. This part of the exposition draws on Friston (2009, 2013), Friston et al. (2012), and particularly Parr et al. (2022). Interested readers will find sources such as Clark (2013, 2016), Hohwy (2016), Barrett (2017), Otten et al. (2017), and Weise and Metzinger (2017) to be useful entrées to the key ideas. With this review as foundation, we address our ultimate aim, which is to consider how the Bayesian brain may evince an understanding of the processes engaged when a person sits with an assessor to participate in the inkblot task developed by Hermann Rorschach.

The Bayesian Brain

The Broad View

All living organisms face the same challenges: to maintain separateness or boundedness from their surrounding environment. Maintaining separateness is essential to permit efforts towards sustenance and reproduction, as without separateness organisms would dissolve into their surroundings by the forces of entropy (i.e., natural disorder, disorganization, dissipation, and death), per the second law of thermodynamics. Further, all organisms evolved to persist in, or exploit, a relatively narrow band

---

[1] Thomas Bayes (~1701-1761) was a British minister and statistician who developed a relatively simple equation to convert one's current belief about an outcome or event (E; called the prior probability) into a revised and updated belief (called the posterior probability) after encountering some new piece of information, which can be considered a sensory sign or signal (S). The result (i.e., the posterior probability) is a conditional probability because it depends on (i.e., is conditioned on) the new piece of information (i.e., it is E given S, or in symbols E | S). Although most resources present the calculations for Bayes theorem using proportions, Bayes did not and the math is simpler to understand using frequencies (e.g., Gigerenzer & Hoffrage, 1995). To compute the updated conditional probability, one needs to know how often the outcome or event occurs in the presence of the signal (E&S) and how often the signal shows up naturally (S). With those two pieces of information, the posterior probability is simply E&S / S. For instance, say it rains (E) on cloudy days (S) 3 times per month (i.e., E&S = 3) and it is cloudy 9 days per month (i.e., S = 9). On any given day, one's confidence it will rain that day (i.e., prior probability) is roughly 10% (i.e., 3 / ~30). However, if it is cloudy that day, one's confidence it will rain (the posterior probability, E | S) is 33% (i.e., 3 / 9).

of preferred physiological or characteristic states out of all those available, so that all organisms must also avoid those states that are poorly suited to their existential needs (e.g., fish need to live in water, but also need to avoid waters beyond their preferred temperature range and salinity).

According to the model on offer, organisms fulfil their particular aims by actively minimizing the likelihood of encountering environmental circumstances that generate unexpected sensory states, which are conceptualized as ‘surprising’ sensory observations. Surprise in this context has a technical meaning (in information theory it is called self information), in that it indexes the extent to which the current sensory input differs from expected or preferred sensory inputs that are conducive to the persistence of the organism. Because it is generally impossible for an organism to know the true causes of its sensations, minimizing surprise itself is an intractable problem; so instead, organisms minimize an approximation to it called free energy. Free energy, like surprise, is an index of entropic mismatch between an organism’s preferred states and its current sensory samples of the environment. Crucially, free energy is determined just by the organism’s expected or preferred states and their sensory experience of the environment at the moment, which, of course, is just a proxy for the actual environmental causes of that sensation.

Separateness and boundedness are not just requirements of organisms, but also are required for any organized, adaptive system. Within an organism there may be many organized systems (e.g., a brain encased within a skull, a vascular system throughout the body). For each, there is a boundary individuating it from its surroundings. What is on the other side of that boundary is hidden from that within the boundary (e.g., the external environment is inaccessible for a brain in a skull). Statistically, the boundary functions as what is known as a Markov blanket, which mediates between outer states and inner states; namely states that are external to the system and those which are internal to it. From the perspective of active inference, the internal system has only two options for inferring what is outside of the system, both of which are mediated via the blanket; the blanket can mediate action from within to without (via active states such as muscles and autonomic reflexes) and it can sample the environment (via sensory states such as sensory epithelia and receptors) from without to within. However, the internal states of the system cannot directly influence the outer states and vice versa; inner influences outer only through active states (e.g., actions) and outer influences inner only through sensory states (e.g., observations).[2] Together, the active and sensory states constitute blanket states.

In any adaptive system capable of persisting, active states and sensory states have a reciprocal or symmetrical relationship with each other, in that active states influence outer states and outer states respond by providing updated sensory states for the organism. This permits the organism (the inner state) to update conditional Bayesian probabilities to infer its external states and thus make probabilistic inferences about the causes of its sensations (e.g., a finger stroke on an object generates a sensation of roughness, increasing the probability of it being a certain type of hidden cause and lowering the probability of it being others). Those internal inferences update Bayesian prior probabilities or initial expectations that then form the baseline for what to expect from the environment or niche as the organism acts to sample it further. Those inferences thus serve to foster what can be understood as a model or representation of the external environment, reducing its uncertainty and implicitly, actively reducing the unpredictability of the external states of affairs in the niche.

In short, this reduction helps minimize free energy, which is an index of unexpected environmental surprises that generally counter adaptive persistence. For instance, a fish that swims left yet senses a large shadow moving similarly overhead may associate that sensation with an increased likelihood (i.e., probability) of danger, which should increase in certainty further if it then swims away

---

[2] The term blanket comes from the idea of enclosing or enveloping. In human terms, a Markov blanket would be one's outer surface, which is largely skin, though the term skin is generally not used for the inner parts of the mouth, nose, or ears. To know what resides outside oneself, the action options include to look, listen, sniff, taste, touch, move (ambulate), or think. The options to sense outside the body are mainly through the eyes, ears, nose, mouth, and skin, using sensory mechanisms that have an adaptively limited range of functions to detect what it truly external to the self (e.g., people cannot smell like dogs, hear like bats, or use magnetism like birds).

and the shadow follows. Thus, although they vary tremendously in their complexity, adaptive systems form a type of 'understanding' or recognition of their environment through their bounded exchange with that environment. From an active inference perspective, this type of recognition unfolds via process of Bayesian belief updating, under a generative model.

Generative models do more than track linked patterns of active and sensory states. They serve both as a probabilistic base of inference (i.e., Bayesian prior probability) and as a mechanism to advance the needs of the inner state in order for the organism to persist over time. Further, the generative model encompasses the intrinsic needs of the organism; in the sense that those intrinsically preferred, adaptive states form a core base set of prior probabilities for that organism. These core prior probabilities guide the organism to find itself in just that niche, such that it literally perceives itself to be in—and acts to get in—environmental states that are conducive to its survival. Doing so reduces the mismatch between the organism's predictions generated by its internal model and its sensations (i.e., it minimizes free energy).

A generative model can be quite simple, or it can be organized into greater degrees of complexity, with multiple factors, modules, modalities and domains; each serving distinct subsystems for the organism (e.g., temperature regulation versus response to threat). In addition, increasingly complex generative models are hierarchically organized to permit organisms to make inferences across different timescales and to subserve distinct preferences or expected states. Systems with the capacity to model alternative courses of action and their predicted consequences, and to correspondingly be able to engage their environments with agency, likely form the foundation for all sentient organisms, including people. These systems need not just minimize free energy in the moment (*variational* free energy), but also minimize the free energy associated with different possible courses of action in the future (*expected* free energy).[3]

All life forms—and any adaptive system that actively samples its sensory states to minimize variational free energy—entails a generative model. In this formulation a generative model is simply a probabilistic representation that can guide some form of action, including something as simple as secreting a chemical. Such systems can retain their boundaries, regulate their internal states, and persist in an environment suited to their particular life needs. However, the distinction between variational and expected free energy offers a 'bright line' between simpler life forms, which exist solely in the present (even if they have future oriented adaptations, such as trees that drop their leaves annually), and more complex life forms that can explicitly plan and select among possible alternative futures. The latter permits at least some level of deliberation, agency, and sentience.

The Focused View

Active inference in sentient organisms operates using hierarchical Bayesian predictive coding, where specialized neuronal paths send predictions of what to expect from higher cortical levels via representation unit neurons to the lower levels, ultimately reaching the sensory epithelia. In this hierarchical or deep architecture, each level conveys predictions to the next level down. Return signals from the sense receptors then traverse each layer of the hierarchy, but do so using error unit neurons. These neurons convey prediction errors or 'surprising' information that is not explained by the downward flowing predictions. These prediction errors inform top-down cells at their level and at the level directly above. (This process of returning residuals [i.e., predicted experience – encountered experience = prediction error] rather than already known information is efficient metabolically and somewhat analogous to lossless file compression techniques, which discard predictable information

---

[3] The distinction here is between something like determining if you should put another layer of clothing on now to counter an emerging chill (minimize variational free energy) and deciding if it is worth the effort to carry an umbrella to counter a potential chill from rain later (minimizing expected free energy).

[e.g., .png or .zip file formats]). Importantly, predictions or prior beliefs (i.e., Bayesian prior probabilities) are themselves probability distributions, not point estimates, meaning for every representation an organism has—of the hidden environmental causes of their sensations—there is an expectation (i.e., mean) and an associated degree of confidence or precision (i.e., variance) that determines the degree of certainty with which they are held. High confidence produces a narrow, tight distribution, while low confidence a broad and diffuse distribution. Similarly, sensory inputs, and the degree of prediction error they return to trigger a revision of predictions (i.e., Bayesian posterior probabilities) or affirm a match with prior predictions, also are probabilistic, with their magnitude (mean) and precision (variance) determined by the quality and reliability of the sensory signal. In short, each level in the hierarchy encodes the uncertainty or precision associated with its current (Bayesian) beliefs that affects both the top down predictions and the bottom up errors. This precision is thought to be mediated by the frequency and intensity of the synaptic signals at each juncture. More specifically, precision may be encoded by the sensitivity of prediction error units to their inputs; such that a high degree of precision at one level of the hierarchy means that the prediction errors have a greater influence on belief updating at superior levels. Physiologically, this corresponds to synaptic gain, while psychologically, it can be regarded as an implementation of attentional gain—or its attenuation.

Each level of this hierarchical structure seeks to reconcile or cancel or resolve prediction errors to reduce uncertainty and optimize its representation of external states. Note that minimizing prediction errors is just the same as minimizing free energy or surprise in our treatment of adaptive exchange with the environment above. Importantly, agents can amplify certain prediction errors by focusing attention on the signals provided, or they can inhibit that input to attenuate messages. For instance, sensory attenuation is required for an agent to take any action, because any effort to move would be countered otherwise by error signals conveying that in fact that movement is not taking place. The constant interplay of the predictions and prediction errors, up and down the neural hierarchy, leaves considerable room for problems to develop in the sensorium, such that prior beliefs may be inappropriately amplified or attenuated (higher or lower prior precision) or held too tentatively or confidently (greater and lesser prior variance). The same is true with respect to sensory input, which can be inappropriately strong or weak (higher or lower sensory precision) or dismissed or amplified (wide or narrow distribution).[4]

At the same time, it is quite adaptive, in the right context, to give excess weight to one level of hierarchal processing or another. Highly weighted sensory signals can quickly update one's (Bayesian or subpersonal) beliefs about the environment, which can be helpful when in novel or dangerous surroundings, changing or chaotic circumstances, or if the signal is particularly strong and precise. Highly weighted prior predictions are stubbornly resistant to change and influence, which can be helpful when operating under stable and familiar conditions or when the environmental signals are vague, contradictory, or confusing.

Thus, the active inference model, in its broadest form, posits that all creatures seek to find and make niches over momentary and lifelong time frames that minimize prediction error about the hidden external causes of sensation, as that error is a manifestation of the deleterious and dissipative forces of entropy (i.e., unpredictability is a manifestation of disorder or randomness in the life trajectory of the organism). They do so by constantly predicting what the agent should experience and sense and constantly affirming or updating those predictions based on the error notifications from the senses.

Active inference itself is the process of resolving discrepancies between one's model of experience (i.e., prior predictions) and the sensations generated by actual but hidden environmental signals. One can do this either by taking some action (e.g., allocate attention, step back, shift gaze, smell more deeply, turn an ear, touch again) to resample sensory information to confirm the prediction or by

---

[4] For instance, one could feel more or less hungry than is true physiologically (incorrectly high or low signals) and one could be uncertain or very confident about that level of hunger (incorrectly wide or narrow signals).

letting error signals override and correct the generative model of prediction. Changing one’s mind to accommodate updated beliefs is *perception*. Modifying the environment sensed or sampled is *action*. Active inference thus leads either to *affirmation* of the prediction via action bringing affirming sensory samples with minimal prediction error or to *correction* to foster a refined prediction that resolves prediction errors via perception— and that is now accompanied by new recognition. Both paths serve the aim of minimizing the discrepancy between one’s model of the environment and the environment itself (i.e., minimize free energy). It is easy to find illustrations of these ideas through common expressions that often emerge in conversation using terminology for embodied cognition. For instance, a listener may exclaim, “Oh, I see!” as a manifestation of perception (i.e., arriving at a new corrected understanding or view) or a speaker may offer guidance for action by saying, “No, look at it this way; …” or “Consider it from this perspective; …” directing the listener to mentally move to a different position to see their point.

The Conventional View

Although there is much more that could be said about active inference and the Bayesian brain, this overview sketches out its main features. The view that the brain is fundamentally a prediction machine dates back to Helmholtz in the 1860s. However, it is worth highlighting how this model of minds, brains, and nervous systems is in contrast to the alternative view that has been conventional for decades, which is of sensory cells taking in specific features of the environment and using the neural hierarchy to build an increasingly complex and accurate understanding of what is being perceived at higher cortical levels (e.g., Aggelopoulos, 2015). Rather than viewing perception as a one-way process of passively taking in sensory information and then trying to figure it out, the actively inferring Bayesian brain begins with a generative model that uses an active, constructive process going from the inside out, such that sensations of and from the environment—which are shaped, shifted, and refined by the agent's actions—serve to affirm or modify the organism's model of the hidden causes behind those sensations. Complex agents infer the source of their experiential sensations, with experiential sensation encompassing the world of objects and actions outside oneself and the similarly remote or hidden world of processes, impulses, affects, and needs inside one's own body.

Empirical Grounding and Applications of the Model

The active inference model encompasses many facets. However, they all are grounded in rigorously defined mathematical formalisms that are tightly linked to biologically plausible mechanisms of action in the context of evolutionary developments for all living organisms, sentient or not. The paradigm has been remarkably heuristic (Clark, 2016; Parr et al., 2022), providing guidelines for more advanced robotics and artificial intelligence, as well as explaining brain localization and functions and their proclivity for using specific neurotransmitters. Within psychological concerns more specifically, these models are being used to conceptualize topics as diverse as psychedelic experiences (Carhart-Harris, 2018), hypnosis (Martin & Pacherie, 2019), interoception (e.g., Seth, 2013), attention (e.g., Parr & Friston, 2017), trauma (e.g., Linson et al., 2020), schizophrenia (e.g., Friston et al., 2016), hallucinations (e.g., Benrimoh et al., 2018), delusions (e.g., Adams et al., 2014), autism (e.g., Palmer et al., 2017), movement disorders (e.g., Brown et al., 2013), Freud's unfinished *Project for a Scientific Psychology* (Carhart-Harris & Friston, 2010), and consciousness itself (Solms & Friston, 2018), to name a few.

Links to Engaging with the Rorschach Task

With this as backdrop, we turn to consider several ways in which active inference can be applied to Rorschach's inkblot task. In particular, we attempt to make links with some of Rorschach's own views of what the task was, how it worked, and, in consequence, what it provided. For those views, we rely heavily on the excellent and substantially clearer new English translation of his work by Keddy, Signer, Erdberg, and Schneider-Stocking (Rorschach, 2021).

Key Features of the Task

As readers of this journal know, the Rorschach is a performance task that relies on visual-spatial and lexical-conceptual problem solving, using the standard set of 10 vertically symmetrical inkblot designs set on white cardstock. Five are shades of gray, two are shades of gray with prominent bold red areas, and three are fully chromatic with elements ranging from pleasing pastels to brightly saturated colors. For administration, the assessor hands respondents each card in order, while asking the question "What might this be?" Respondents typically reply with two or three responses per card and their replies represent their personal solutions to deciphering the problem at hand. Subsequently, the assessor goes back through the cards and clarifies where objects reside and what inkblot features contributed to that perception. Finally, most assessors then classify each response across multiple dimensions (e.g., use of color, envisioning human activity, coherence of thought processes), and aggregate the codes across all responses to form summary scores capable of contrasting the respondent to what most people see, say, and do when completing the task.

Although likely less well known to many readers, the inkblots are not random designs, despite Rorschach referring to them in the subtitle to his text as 'accidental forms.' To the contrary, Rorschach used his artistic training to carefully create, pilot-test, and artistically refine each card over time to ensure they would not simply look like inkblots (Rorschach, 2021; Searls, 2017).[5] He appears to have had two intertwined aims when developing them, both based on the suggestive 'critical bits' (Exner, 1996) that encompass the prominent inkblot areas and shapes and also their color, shading, irregular interior and exterior contours, and symmetrical features. First, within the designed composition of each card, he

---

[5] The Rorschach Archives contain multiple iterations of each of the 10 cards that are clearly recognizable but differ from the final version in their accentuation of features and their overall composition. Of this, Rorschach (2021, p. 4) said, "The picture series used in the test gradually developed on the basis of empirical observation."

embedded at least one reasonably recognizable object or part of an object to form the commonly reported conventional percepts (e.g., the human like figures on Card III and the animal like figures on Card VIII). Second, he embedded a textured array of other features that contradict or complicate the more recognizable elements (e.g., a part looks pretty clearly like a person's head, but what would be its torso looks more like the head of a horned creature).

These opposing qualities produce evocative but incomplete or imperfect perceptual likenesses that (deliberately and artfully) stimulate uncertainty, ambiguity and imprecision among the competing visual impressions that may underwrite potential responses. They also provide a task with considerable embedded structure, as well as a wide array of alternative features that idiographically hook perception and contribute to personally unique perceptions. The embedded structure provides a mechanism for assessing conventionality in the locations selected for percepts (i.e., the focus of one's attention; Berry & Meyer, 2019) and the quality of the fit of objects to those locations (i.e., perceptual accuracy as coded by Form Quality). The idiographic diversity provides personally salient, experience-near imagery that can richly illustrate the respondent's psychological processing. Interestingly, even in very large samples, unique objects seen by just a single person account for about 70% of all the distinct objects reported (Meyer et al., 2011).[6]

The task of dealing with imprecision in the provocative and deliberately contradictory stimuli, as well as uncertainty regarding the adequacy of one's responses, occurs in an interpersonal context; while the respondent interacts with a relative stranger (sitting adjacent) who is observing and transcribing the exchange. These features make the task moderately stressful, and more stressful than assessment by self-report methods (e.g., Momenian-Schneider et al., 2009; Newmark et al., 1974, 1975).

---

[6] For instance, Villemor-Amaral and her colleagues (cited in Meyer et al., 2011) identified a total of 6,459 response objects in a sample of 600 nonpatients. The most common object was identified by 375 people and just 30 objects were seen by 50 or more people; however, 4,538 objects were identified by just one person.

For the respondent, solving the problem of what the inkblot might be invokes a series of perceptual and problem-solving (belief updating) processes. These include scanning the stimuli, selecting locations for potential response objects, comparing potential inkblot images to objects in memory, evaluating possible percepts relative to their inconsistencies or contradictions, reformulating response options, filtering out those judged less optimal, and articulating a final solution to the assessor (Exner, 1974). The respondent's visual-mnemonic matching of objects in the card to recalled images, conceptual processing of the stimuli, and verbal and nonverbal communication engage all brain regions, encompassing bilateral activity in the frontal, temporal, parietal, occipital, and limbic lobes (e.g., Asari et al., 2008, 2010a, 2010b; Giromini et al., 2017).

Psychological Processes Engaged

The available neurophysiological data indicate that completing the task engages both the dorsal and ventral attentional systems (Giromini et al., 2017). The dorsal system is important for directing conceptually guided top-down attentional search processes (i.e., predictions of what it might be) and the ventral system is important for recognizing and reorienting to surprising or unexpected bottom-up input (e.g., misfits with prediction, alternative possibilities). These two attentional systems are negatively correlated with the default mode network (e.g., Zhou et al., 2018), which in humans is implicated in self-referential processing, including the introspective attribution of self-reported characteristics (e.g., Davey et al., 2016).

These data on Rorschach responding fit nicely with an active Bayesian brain, as the respondent is iteratively refining the fit of conceptual priors (beliefs about what it might be, carried by the dorsal attention system) to environmental (visual) stimuli with an uncertain or ambiguous cause. Given their intentionally contradictory features, the Rorschach images consistently provide the viewer with irreducible error signals that the prediction is not quite right and is evincing ill-fitting incongruities (ventral attention system). Respondents reduce this prediction error to modify the initial prediction about what is ‘out there’ in the environment (i.e., by changing one’s mind about the percept) or by taking actions to sample the environment (e.g., shift gaze, modify location boundaries) and more precisely affirm the prediction (i.e., by gathering better evidence). This results in an iterative calibration process that ultimately provides the respondent’s error-corrected, personalized perceptual equilibrium (i.e., beliefs) about the hidden features of the environment. Of course, here the inkblots are serving as analogs to the parallel processes occurring when encountering the many ambiguities of daily life (Clark, 2016).

In other words, the Rorschach task presents a carefully designed and special problem for perceptual inference; in that there is no single perceptual explanation that fully accounts for the visual information at hand. This provides an unprecedented tool to explore the landscape of a subject’s prior beliefs about the causes of their sensations. In one sense, the Rorschach’s task is the ultimate tool for disclosing prior beliefs. It is reminiscent of how psychophysics reveals prior beliefs through the use of illusions: illusory stimuli (e.g., ambiguous figures or stimuli that induce bistable perception) are carefully constructed to induce ambiguity; obliging the perceiver to explore and alternate between perceptual hypotheses that reveal the kind of hypotheses people use in everyday perceptual synthesis.

This ambiguity is not an accidental feature of Rorschach’s task. He designed it to engage just this form of active inference to allow an assessor to see the meaning-making process in action. Rorschach (2021) viewed the task as one of “perception and apperception” (p. 36), not imagination or gaining access to unconscious processes per se. He considered *perception* to be the outcome of an associative process between sensation and one’s memory of former sensations, paralleling how that term is used in active inference. *Apperception* was the process of linking sensory perceptions with their prior connections in order to understand current sensations based on past experience. This can be viewed as the process of getting to perception, which from an active inference account is the process of balancing predictions and errors to ‘know’ what is experienced. However, Rorschach (2021) saw one big difference between those process in everyday life and those processes when examining his inkblots; in essence, the inkblots slowed perception to render it visible.

> If perception [is] an associative *integration of present engrams* [memory traces] *with recent complexes of sensations*, then *the interpretation of accidental* [indeterminate] *forms can be called a perception in which the effort of integration of the sensations and the memory trace is so great that it is perceived as an effort of integration*. This intrapsychic perception of incomplete equality [i.e., discrepancy] between the complexes of sensation and the engram gives the perception the character of an interpretation. … Most respondents with either schizophrenia, epilepsy, manic-depressive illness, or organic disorders… are not aware of the effort of integration. Even many normal respondents are not aware of it. These respondents do not interpret the pictures, they name them. They may even be astonished if other respondents see something different in them. In these cases, this is not an interpretation but rather a perception in the strict sense of the word. They are as unaware of the associative effort of integration as a normal person is when recognizing a familiar face or perceiving a tree. Therefore, there must be a kind of threshold beyond which perception (the integration without awareness of the effort of it) becomes interpretation (perception with awareness of the effort of integration). … In summary, we may conclude that *the differences between perception and interpretation are based on individual and gradational factors, not on general, principal ones; thus, interpretation may be a special case of perception.* There is, therefore, no doubt that the form interpretation experiment [i.e., the inkblot task] can be called an investigation into perception. (pp. 36-37, italics in the original)

Thus, Rorschach recognized that the task, while one of perception, also helped show the iterative cycle of prediction, error correction, active search, revised prediction, further error correction, further action, and so on. Of course, this view is quite compatible with the iterative processes undertaken by the Bayesian brain. However, Rorschach (2021) provides further elaboration of his views on what the task does and does not provide later in the text, when discussing interpretation. He considered the extent to which respondents mentally enlivened their perceptions with human activity and the extent to which they recognized and incorporated the bright, vibrant coloration of the cards as key dimensions that differentiated individuals. However, he was clear that there was not a direct correspondence between the nature of one's perceptions and behavior in everyday life. Thus, after quantifying these two dimensions Rorschach concluded that the assessor would "know a lot about the respondent" (p. 106). He further clarified that what the task revealed was the sensory-perceptual structures of the individual that registers their day to day experiences; revealing their lived experience but not the way they live their life.

> We do not know what this respondent experiences, but, rather, *how this respondent experiences*. We know a large part of the characteristics and dispositions with which the respondent goes through life, be they of an associative or affective nature or a mixture of the two. We do not know their experiences, but we do know the *experience apparatus* [also psychical apparatus] with which they receive experiences from the inside and from the outside, and to which the respondents initially subject their experiences to processing. (p. 106, italics in the original)
>
> The experience apparatus with which the individual experiences is a much broader, more extensive structure than the apparatus with which the individual lives. To experience, a person has a number of registers but only uses a few for the actions of life – often so few that it ends in

> stereotypy. The experience type [balance of movement to color] reveals how broad the apparatus is with which the respondent could live. It cannot reveal, however, actually – except under very favorable circumstances – what parts of the apparatus the respondent activates for active living. (p. 108)

Envisioning Human Action

Another domain in which the actively inferring Bayesian brain manifests in Rorschach responding is with respect to envisioning human activity. The mirror neuron system is activated when a person engages in a particular course of action and similarly activated when observing another person engaging in that particular course of action. It is viewed as the neuronal representation of understanding what others are doing (proprioceptive and exteroceptive) and why (goals and intentions) when we observe their behavior and actions in a particular context (e.g., Friston et al., 2011; Kilner et al., 2007). In essence, we use our experience of a movement or activity in context to understand another's movement or activity in that context, and this is done through active inference, mentally anticipating (predicting) the act and modifying the mental prediction based on mismatches (errors) between the observed and internally enacted action. On this basis, one could anticipate how seeing human activity in inkblot imagery would similarly activate the mirror neurons.

Indeed, Rorschach (2021) anticipated as much when describing the human Movement code, M. "Movement responses (M) are those interpretations which are determined by *form perception plus kinesthetic factors*. The respondent imagines the object interpreted to be in motion" (p. 45, italics in the original). Thus, Rorschach was identifying an empathic, enactive internal response to the action perceived. Subsequent neurophysiological research using multiple methods (EEG, TMS, MRI) has affirmed these views, with clear evidentiary support that the mirror neuron system is activated when respondents produce responses coded M, but not when simply seeing static human figures or animals in action (see e.g., Giromini et al., 2019).

Introversiveness and Motility: An Open Question

Rorschach was fascinated with movement, including its artistic depiction, its consequences for mental life, its manifestations in culture and cults, and, of course, in responses to his inkblots (Akavia, 2013). For the latter, he used his considerable skills depicting action in drawings and paintings to provide movement-suggestive stimuli in the inkblots in order to understand the type of person who responded to it. Rorschach (2021) believed M responses required a degree of delay and reflection to formulate, reflecting a style of processing that was ideational and introversive (i.e., “capable of introversion”; p. 97). In contrast to the zeitgeist at the time (Akavia), he also believed there was an inverse relationship between physical movement and perceiving human movement in the inkblots.

> “*The measure of the manifest motility in a respondent is not the measure of the kinesthesias* [responses with M] influencing the respondent during the perception process. *On the contrary, the kinesthetic individual is motorically stable; the lively person is poor in kinesthesias.*” (p. 45, italics in the original)
>
> And further, “Introversion … is increased by an active shutting down of the factors that inhibit it and decreased by restarting the function of adaptation” (p. 97).

Indeed, among patients with schizophrenia, Rorschach (2021) concluded that those with catatonia produced the highest number of responses coded M (Table 2, p. 42). Rorschach had extensive experience with patients who had schizophrenia and he wrote a lengthy unpublished case study of Theodor Niehans, a patient he assessed and treated at the Münsingen asylum, with a well-documented

20-year history at the asylum (Akavia, 2013). Over the years, Niehans went from profound paranoia to profound catatonia and back again. Akavia notes the following, “Rorschach ultimately conceptualized the catatonic form of schizophrenia, ostensibly a state of extreme stasis, as manifesting an intense internal dynamism of ‘fettered movement’” (p. 6), which Niehans himself characterized as a period of “compulsive thought” (p. 121). In the case study, Rorschach contrasted paranoia and catatonia, saying, “the catatonic renounced the outside world and abandoned himself to introversion, while the paranoid resisted introversion by desperately cleaving to the outside world” (p. 106). Akavia concluded “Rorschach… saw catatonia and its concomitant physical immobility as a mode of utmost mental excitability, whereby—at least in the case of Niehans—schizophrenic psychological activity found its consummate inward form, giving rise to an active, ongoing development of florid delusions.” (p. 123).

Interestingly, Brown et al. (2013) used active inference to model the consequences of a compromised ability in the typical requirement to attenuate sensory signals during self-generated movement. This attenuation is required to initiate action and it is commonly compromised in patients with schizophrenia. Failing to attenuate those sensory signals leaves sensory signals stronger and more precise than one’s predictions of movement; as such, sensory prediction errors predominated over top-down projections. Under these conditions, they observed profound impairment of movement, reminiscent of the psychomotor symptoms of catatonia. Although their modeling did not encompass ideational activity, it is a fascinating open question about whether these conditions would lead to an increased number of ideationally active human movement responses on the inkblot task, as Rorschach’s observations suggest.

Rorschach’s (2021) notion that inkblots ‘slowed perception to render it visible’ is exactly congruent with the definitive role of precision or uncertainty coding in the covert action associated with sensory attention and attenuation. This follows because the precision determines the rate of belief updating. In other words, assigning a greater weight to certain prediction errors means they have a greater influence on neural populations encoding expectations and subsequent predictions higher in the hierarchy. Precluding a precise, high level prior explanation for sensory input will therefore preclude precise prediction errors higher in the visual (or more generally perceptual) hierarchies—and thereby attenuate the rate of belief updating or assimilation of prediction errors from lower levels. This corresponds exactly with the notion of slowing perceptual synthesis so that its architecture and fundament's can be disclosed through responses reported to the assessor.

Perceptual Distortions

Finally, we consider perceptual distortions, which from a Bayesian brain perspective may emerge from either end of the neural hierarchy, such that hallucinations and perceptual distortions may occur when sensory signals fail to be attenuated or perceptual priors are underweighted (e.g., Adams et al., 2013) or, more typically, when prior beliefs are over-weighted and corrective sensory input is under-weighted during percept formation (Benrimoh et al., 2018; Corlett et al., 2019; Parr et al., 2018). When coding Rorschach responses most contemporary systems for use differentiate several levels to designate the quality with which percepts fit the inkblots at the location being used, known as Form Quality, including conventional or ordinary, unusual or idiosyncratic, and distorted or minus (e.g., Meyer & Mihura, 2021). Rorschach (2021) noted that to produce responses with good form quality, respondents needed stable attention, clarity in their efforts at perceptual and associative integration, and self-control. As such, among his patients with schizophrenia, only those with paranoia produced reasonably conventional responses, while those with disorganized symptoms had a higher frequency of distorted or idiosyncratic perceptions. These observations by Rorschach have received consistent support in the subsequent research literature, indicating Form Quality is an excellent marker of perceptual deviance and one of the best validated measures derived from the Rorschach (Meyer & Mihura, 2021; Mihura et al., 2013).

Rorschach research bearing on these issues that could inform an active inference model of perception support the notion that unique perceptions can be associated with unduly weighted priors and insufficient regard for corrective environmental sensory feedback. Asari and colleagues (2008, 2010a, 2010b) conducted three interrelated studies of Form Quality scores that provide an understanding of the psychological operations active when a respondent is generating a response with particular types of Form Quality. The authors used functional and structural MRI to examine the neurophysiological features associated with atypical and distorted perceptions, uncommon perceptions, and conventional perceptions. A key finding was that amygdala activity in people giving atypical and distorted perceptions generated a positive, excitatory link between the right temporal pole and the left anterior prefrontal cortex, while simultaneously generating a negative, inhibitory effective connectivity from the right temporopolar region to the bilateral occipitotemporal regions.

Thus, atypical and distorted perceptions in this study involved instances when internal processes triggered by something in the inkblots activated affectively charged brain structures to turn off the typical process of reciprocal visual calibration between ideas and perceptual stimuli in favor of idiosyncratic, nonconsensual top-down views. In essence, sensory signals were shut down in favor of overly precise charged beliefs. Rather than taking in the visual cues the environment was providing, personally relevant, emotionally salient unique perceptions forced themselves into an inkblot representation, overriding the respondent's ability to perceive experiences in a conventional manner.

Concluding Comments

With this essay, we hope to have interested readers in the richly productive and increasingly broad literature on active inference and the Bayesian brain. In some depth, we outlined aspects of the evolutionary, biological, and neurophysiological foundation for this mathematically inspired model of functioning. We also identified psychologically relevant areas of active research that readers my find useful for further exploration and we closed by offering a handful of ways that Rorschach responding appears to fit seamlessly and meaningfully within an actively inferring Bayesian brain.

References

Adams, R. A., Brown, H. R., & Friston, K. J. (2014). Bayesian inference, predictive coding and delusions. *AVANT. The Journal of the Philosophical-Interdisciplinary Vanguard*, *V*(3), 51–88. https://doi.org/10.26913/50302014.0112.0004

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. *Frontiers in Psychiatry*, *4*.

Aggelopoulos, N. C. (2015). Perceptual inference. *Neuroscience and Biobehavioral Reviews*, *55*, 375–392. https://doi.org/10.1016/j.neubiorev.2015.05.001

Akavia, N. (2013). *Subjectivity in motion: Life, art, and movement in the work of Hermann Rorschach* (2012-34295-000). Routledge/Taylor & Francis Group.

Asari, T., Konishi, S., Jimura, K., Chikazoe, J., Nakamura, N., & Miyashita, Y. (2008). Right temporopolar activation associated with unique perception. *NeuroImage*, *41*(1), 145–152. https://doi.org/10.1016/j.neuroimage.2008.01.059

Asari, T., Konishi, S., Jimura, K., Chikazoe, J., Nakamura, N., & Miyashita, Y. (2010a). Amygdalar enlargement associated with unique perception. *Cortex*, *46*(1), 94–99. https://doi.org/10.1016/j.cortex.2008.08.001

Asari, T., Konishi, S., Jimura, K., Chikazoe, J., Nakamura, N., & Miyashita, Y. (2010b). Amygdalar modulation of frontotemporal connectivity during the inkblot test. *Psychiatry Research: Neuroimaging*, *182*(2), 103–110. https://doi.org/10.1016/j.pscychresns.2010.01.002

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(11), 1833. https://doi.org/10.1093/scan/nsx060

Benrimoh, D., Parr, T., Vincent, P., Adams, R. A., & Friston, K. (2018). Active inference and auditory hallucinations. *Computational Psychiatry (Cambridge, Mass.)*, *2*, 183–204. https://doi.org/10.1162/cpsy_a_00022

Berry, B. A., & Meyer, G. J. (2019). Contemporary data on the location of response objects in Rorschach's inkblots. *Journal of Personality Assessment*, *101*(4), 402–413. https://doi.org/10.1080/00223891.2017.1408016

Brown, H., Adams, R. A., Parees, I., Edwards, M., & Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing*, *14*(4), 411–427. https://doi.org/10.1007/s10339-013-0571-3

Carhart-Harris, R. L. (2018). The entropic brain—Revisited. *Neuropharmacology*, *142*, 167–178. https://doi.org/10.1016/j.neuropharm.2018.03.010

Carhart-Harris, R. L., & Friston, K. J. (2010). The default-mode, ego-functions and free-energy: A neurobiological account of Freudian ideas. *Brain: A Journal of Neurology*, *133*(4), 1265–1283. https://doi.org/10.1093/brain/awq010

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K., & Powers, A. R. I. (2019). Hallucinations and strong priors. *Trends in Cognitive Sciences*, *23*(2), 114–127. https://doi.org/10.1016/j.tics.2018.12.001

Davey, C. G., Pujol, J., & Harrison, B. J. (2016). Mapping the self in the brain's default mode network. *NeuroImage*, *132*, 390–397. https://doi.org/10.1016/j.neuroimage.2016.02.022

Exner, J. E. (1974). *The Rorschach: A Comprehensive System* (1st edition, Vol. 1). Wiley.

Exner, J. E. (1996). Critical bits and the Rorschach response process. *Journal of Personality Assessment*, *67*(3), 464–477. https://doi.org/10.1207/s15327752jpa6703_3

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K. (2013). Life as we know it. *Journal of The Royal Society Interface*, *10*(86), 20130475. https://doi.org/10.1098/rsif.2013.0475

Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, *3*, Article 151. https://doi.org/10.3389/fpsyg.2012.00151

Friston, K., Brown, H. R., Siemerkus, J., & Stephan, K. E. (2016). The dysconnection hypothesis (2016). *Schizophrenia Research*, *176*(2–3), 83–94. https://doi.org/10.1016/j.schres.2016.07.014

Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, *104*(1–2), 137–160. https://doi.org/10.1007/s00422-011-0424-z

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704. https://doi.org/10.1037/0033-295X.102.4.684

Giromini, L., Viglione, D. J. Jr., Pineda, J. A., Porcelli, P., Hubbard, D., Zennaro, A., & Cauda, F. (2019). Human movement responses to the Rorschach and mirroring activity: An fMRI study. *Assessment*, *26*(1), 56–69. https://doi.org/10.1177/1073191117731813

Giromini, L., Viglione, D. J., Zennaro, A., & Cauda, F. (2017). Neural activity during production of Rorschach responses: An fMRI study. *Psychiatry Research: Neuroimaging*, *262*, 25–31. https://doi.org/10.1016/j.pscychresns.2017.02.001

Hohwy, J. (2016). The self-evidencing brain. *Noûs*, *50*(2), 259–285. https://doi.org/10.1111/nous.12062

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, *8*(3), 159–166. https://doi.org/10.1007/s10339-007-0170-2

Linson, A., Parr, T., & Friston, K. J. (2020). Active inference, stressors, and psychological trauma: A neuroethological model of (mal)adaptive explore-exploit dynamics in ecological context. *Behavioural Brain Research*, *380*. https://doi.org/10.1016/j.bbr.2019.112421

Martin, J.-R., & Pacherie, E. (2019). Alterations of agency in hypnosis: A new predictive coding model. *Psychological Review*, *126*(1), 133–152. https://doi.org/10.1037/rev0000134

Meyer, G. J., & Mihura, J. L. (2021). Rorschach Performance Assessment System (R-PAS) for assessing disordered thought and perception. In I. B. Weiner & J. H. Kleiger (Eds.), *Psychological assessment of disordered thinking and perception.* (2021-59355-010; pp. 151–168). American Psychological Association. https://doi.org/10.1037/0000245-010

Meyer, G. J., Viglione, D. J., Mihura, J. L., Erard, R. E., & Erdberg, P. (2011). *Rorschach Performance Assessment System: Administration, coding, interpretation, and technical manual*. Rorschach Performance Assessment System LLC.

Mihura, J. L., Meyer, G. J., Dumitrascu, N., & Bombel, G. (2013). The validity of individual Rorschach variables: Systematic reviews and meta-analyses of the Comprehensive System. *Psychological Bulletin*, *139*(3), 548–605. https://doi.org/10.1037/a0029406

Momenian-Schneider, S. H., Brabender, V. M., & Nath, S. R. (2009). Psychophysiological reactions to the response phase of the Rorschach and 16PF. *Journal of Personality Assessment*, *91*(5), 494–496. https://doi.org/10.1080/00223890903088727

Newmark, C. S., Hetzel, W., & Frerking, R. A. (1974). The effects of personality tests on state and trait anxiety. *Journal of Personality Assessment*, *38*(1), 17–20. https://doi.org/10.1080/00223891.1974.10119934

Newmark, C. S., Wheeler, D., Newmark, L., & Stabler, B. (1975). Test-induced anxiety with children. *Journal of Personality Assessment*, *39*(4), 409–413. https://doi.org/10.1207/s15327752jpa3904_15

Otten, M., Seth, A. K., & Pinto, Y. (2017). A social Bayesian brain: How social knowledge can shape visual perception. *Brain and Cognition*, *112*, 69–77. https://doi.org/10.1016/j.bandc.2016.05.002

Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Bayesian approaches to autism: Towards volatility, action, and behavior. *Psychological Bulletin*, *143*(5), 521–542. https://doi.org/10.1037/bul0000097

Parr, T., Benrimoh, D. A., Vincent, P., & Friston, K. J. (2018). Precision and false perceptual inference. *Frontiers in Integrative Neuroscience*, *12*. https://doi.org/10.3389/fnint.2018.00039

Parr, T., & Friston, K. J. (2017). The active construction of the visual world. *Neuropsychologia*, *104*, 92–101. https://doi.org/10.1016/j.neuropsychologia.2017.08.003

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press.

Rorschach, H. (2021). *Hermann Rorschach's psychodiagnostics: Newly translated and annotated 100th anniversary edition* (P. J. Keddy, R. Signer, P. Erdberg, & A. Schneider-Stocking, Eds. & Trans.; Original publication 1921). Hogrefe Publishing.

Searls, D. (2017). *The inkblots: Hermann Rorschach, his iconic test, and the power of seeing*. Crown.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573. https://doi.org/10.1016/j.tics.2013.09.007

Solms, M., & Friston, K. (2018). How and why consciousness arises: Some considerations from physics and physiology. *Journal of Consciousness Studies*, *25*(5–6), 202–238.

Wiese, W., & Metzinger, T. K. (2017). Vanilla PP for philosophers: A primer on predictive processing. *Philosophy and Predictive Processing: 1.* https://doi.org/10.15502/9783958573024

Zhou, Y., Friston, K. J., Zeidman, P., Chen, J., Li, S., & Razi, A. (2018). The hierarchical organization of the default, dorsal attention and salience networks in adolescents and young adults. *Cerebral Cortex*, *28*(2), 726–737. https://doi.org/10.1093/cercor/bhx307