How Beliefs are like Colors
Devin Sanchez Curry
Synthese 199:3–4 (2021), 7889–7918

Abstract: *Double dissociations between perceivable colors and physical properties of colored objects have led many philosophers to endorse relationalist accounts of color. I argue that there are analogous double dissociations between* attitudes *of belief—the beliefs that people attribute to each other in everyday life— and intrinsic* cognitive states *of belief—the beliefs that some cognitive scientists posit as cogs in cognitive systems—pitched at every level of psychological explanation. These dissociations provide good reason to refrain from conflating attitudes of belief with intrinsic cognitive states of belief. I suggest that interpretivism provides an attractive account of the former (insofar as they are not conflated with the latter). Like colors, attitudes of belief evolved to be ecological signifiers, not cogs in cognitive systems.*

1. Introduction

Recent work has cast renewed doubt on the widespread assumption that ordinary folks construe beliefs as inner causes that produce behavior (Hutto 2011; Tanney 2013; Strand & Lizardo 2015; Dewhurst 2017; Curry 2018; Almagro Holgado & Fernández Castro 2020; Poslajko 2020). This doubt opens the door for a working distinction between what I will call 'attitudes of belief'—the beliefs that lay people attribute to each other (and other animals) in everyday life— and 'cognitive states of belief'—the beliefs that (some) cognitive scientists posit as cogs in cognitive machines.[1] In this article, I will exploit an analogy between belief and color to present a cascading series of arguments for the conclusion that theorists ought not conflate attitudes of belief with cognitive states of belief (pitched at neurophysiological, subpersonal, personal, or etiological levels of psychological explanation). I will also suggest that refraining from conflating attitudes of belief with cognitive states of belief should lead previously reticent philosophers— including identity theorists, psychofunctionalists, commonsense functionalists, etiological teleofunctionalists, eliminativists, and agnostics about the nature of cognitive states of belief—to seriously consider dispositionalism (Ryle 1949; Baker 1995; Schwitzgebel 2002) and especially interpretivism (Dennett 1998; Davidson 2001; Mölder 2010; Curry 2020) about the nature of attitudes of belief.

---

[1] I use the term of art 'attitudes' broadly, following the tradition of Donald Davidson (1963) and Eric Schwitzgebel, to mean "a temporary or habitual posture of the mind" (Schwitzgebel 2013: 76) that is attributed in folk psychological practices. I do not mean to imply by the use of the term 'attitude' that beliefs are attitudes towards propositions; I'm more sympathetic with the claim that they are attitudes towards the world (Sommers 2009).

I use the term 'cognitive', in 'cognitive states', similarly broadly, to mean 'relating to cognition', where cognition is assumed to be constituted by mental (or neural) states (or processes) that cognitive scientists describe in terms of how they contribute to the functioning of the organism who possesses them. Thus, cognitive states (theoretically) occur at neurophysiological, subpersonal, personal, and etiological levels of psychological explanation, since cognitive scientists describe mental/neural states/processes as contributing to the functioning of organisms at each of these levels of explanation.

In §2, I will review the case that colors are neither illusory nor intrinsic physical or dispositional properties of objects. (By the term 'intrinsic' I just mean 'ontologically non-relational'—that is, existing independently of their relation to external perceivers or interpreters.) In §3, I will argue that attitudes of belief are, likewise, neither figments of the folk psychological imagination nor intrinsic physical or dispositional properties of objects. Instead, attitudes of belief are properties organisms have partly in virtue of inhabiting the environments of belief attributors. As such, they ought not be conflated with intrinsic cognitive states of belief—properties organisms are purported to have solely in virtue of their cognitive architecture. In §4, I will argue that insofar as teleofunctionalism helps theorists get a grip on the nature of belief, it should be recruited to support (rather than collapse) the distinction between attitudes and cognitive states. Like colors, attitudes of belief evolved to be ecological signifiers, not cogs in cognitive systems. There would be no believers—no creatures properly characterized as having attitudes of belief— without belief attributors—creatures who characterize creatures as having attitudes of belief. In §5, I will conclude by considering the more radical view that cognitive states of belief, too, emerge relative to belief attributors.

2. Color

Reflection on the metaphysics of color provides an impetus to recognize that attitudes of belief and cognitive states of belief ought to be conceptually distinguished.

2.1. Against conflating perceivable colors with surface spectral reflectance profiles

To see why, consider a well-known problem for the view about perceivable color known as 'physicalism'. Physicalists type-identify perceivable colors with intrinsic properties of colored objects (Smart 1961; Hilbert 1987), just as most philosophers of mind type-identify attitudes of belief with intrinsic cognitive states of belief (Churchland 1981; Fodor 1987; Porot & Mandelbaum 2021). In particular, most physicalists identify colors with either surface spectral reflectance profiles (henceforth SSRs) or the intrinsic physical properties of objects that realize SSRs. Surfaces absorb some light at each wavelength of the visible spectrum and reflect the rest. To have an SSR is to be disposed to reflect a particular percentage of light at each wavelength on the spectrum. Physicalists declare, rightly, that SSRs are the nonrelational dispositional properties of objects most closely associated with their colors. (For example, red objects are disposed to reflect more light in the longer wavelengths than blue objects.) By extension, the physical properties underlying SSRs are the intrinsic physical properties of objects most closely associated with their colors. Nevertheless, identifying colors with SSRs (or their physical realizers) is problematic because there is a double dissociation between the SSRs of objects and the perceived colors of those same objects.

First, perceived colors of objects do not perfectly track SSRs. Cases of metamerism reveal that objects with divergent SSRs often appear identically colored to a single perceiver in a single context. There is no principled way to settle on one metamer's SSR as uniquely veridically captured by the shared perceived color (Hatfield 1992). Likewise, there is no good reason to suppose that only one metamer is *really* the color it appears to be. Thus, perceived colors are multiply realized by SSRs.

Nor does the SSR of an object fix its perceived color. Cases of intrasubjective perceptual

variation reveal that objects with identical SSRs appear differently colored due to context, even when perceived simultaneously by a single perceiver (Cohen 2009: 20). Moreover, the perceived colors of objects vary between perceivers even when the context is held constant. A particularly stark—if atypical—example of this phenomenon known as 'The Dress' went viral on the internet in the 2010s. The Dress is an image of a backlit, striped article of clothing. About 60% of people see The Dress's stripes as alternating white and gold; about 30% see the same stripes as blue and black. Only about 1% of people report being able to (intentionally or unintentionally) switch between the two commonly perceived color configurations (Wallisch 2017: 4). The Dress is thus a clear case of intersubjective variation in perceived color.

Now, this stark difference in how the world is colored for different human perceivers rests on a knife's edge. In most viewing conditions, there is approximate—though not absolute— intersubjective uniformity in perceived color among non-color-blind human perceivers. In normal daylight, the article of clothing photographed for the viral image—'the dress' with a lower-case 'd'—looks blue and black to nearly all humans (including those who see the image of The Dress as white and gold). Importantly, though, there remain subtle differences in the exact shades of blue and black experienced by different individuals (Webster 2015). Due to normal variation in human perceptual systems, even in normal daylight I likely see the dress as either a slightly redder or slightly greener shade of blue than you. And whether the dress's blue is *really* redder or greener is not settled by reference to the dress's SSR.[2]

There is almost certainly more dramatic variance between perceivers of different species. Animals with one or two types of cone cell almost certainly see a differently colored world than trichromatic humans. The same goes for many fellow trichromats; red mason bees detect shorter-wavelength 'ultraviolet' light but not longer-wavelength 'red' light, whereas humans detect 'red' light but not 'ultraviolet' light. Crocuses which look purple to humans likely look yellowish to bees (Menzel et al. 1988).

These cases of intrasubjective, intersubjective, and interspecific perceptual variation reveal that the SSR of an object does not fix its perceived color. In most situations, most perceivers of the same species will have approximately intersubjectively uniform color experiences. However, there are always subtle differences in perceived color, and these differences grow less subtle as perceivers cross species boundaries or enter unusual viewing conditions. Sometimes— as with The Dress, which led to newsworthy amounts of bickering on social media—the differences are stark enough to be practically significant. But the argument for dissociating colors and SSRs need not rely on such dramatic examples (which can be facilely dismissed as illusory). Even with respect to subtle—but ubiquitous—individual differences in color perception, there is no principled way to settle on one perceived color as *the* veridical representation of the SSR in question, much less as *the* true color of the object itself.

---

[2] Empirical research indicates that subjects' expectations about illumination conditions—and attendant modulation of color constancy mechanisms—drive the differential perceived colors The Dress exhibits due to the ambiguous illuminance conditions presented by the backlighting in the image (Wallisch 2017). In normal daylight viewing conditions, there is not much ambiguity in illuminance, so there is not much ambiguity in perceived color. Other researchers have offered (less convincing) explanations of the phenomenon in terms of differential macular pigment optical density (Rabin et al. 2016) or the top-down influence of knowledge of the color of the dress in normal illuminance conditions (Schlaffke 2015).

Philosophers have responded to the double dissociation between perceivable colors and the physical properties of colored objects in several ways. Physicalists and primitivists (Yablo 1995; Gert 2017) insist that each reflective surface has one true perceivable color (or set of colors), and that all others are illusory. For example, physicalists and primitivists might agree that metamers are really distinctly colored (no matter the lighting), that The Dress is really blue and black (even for bees), and that crocus petals are devoid of yellow. Idealists claim that there is no such thing as veridical color perception, because colors are perceptual figments (Jackson 1977). Irrealists claim that there is no such property as color (Hardin 1988; Boghossian & Velleman 1989).

2.2 Color relationalism

Relationalists (Thompson 1995; Hatfield 2009; Cohen 2009; Chirimuuta 2015) hold that colors are neither arbitrary intrinsic properties of objects, subjective sense data, nor illusions. Instead, relationalists respond to the double dissociation between perceived colors and SSRs by arguing that colors are constituted by the relationship between color perceivers and the dispositional properties of environmentally embedded objects. Evan Thompson, for example, writes that "being colored a particular determinate color or shade is equivalent to having a particular spectral reflectance, illuminance, or emittance that looks that color to a particular perceiver in specific viewing conditions" (1995: 245). Thompson thus responds by selectively grounding colors in SSRs that bear the right sort of relationship to color perceivers. In particular, colors are realized by those SSRs that enable perceivers to have color experiences. On this relationalist picture, single SSRs sometimes non-arbitrarily realize different colors *because* they enable different perceivers to see different colors (or single perceivers to see different colors in different contexts). Likewise, different SSRs sometimes non-arbitrarily realize a single color because they both enable perceivers to see that color.

Like Thompson, Gary Hatfield emphasizes the biological and psychological roles of color.

> Not every property is a physical property. The property of being nutritious is not. Neither is color. They are both biofunctional properties. Color, as a property defined in relation to phenomenal experience or psychological discriminatory capacities, is a psychobiological property. As such, its basis may be found in the relation of subjects to objects. (2009: 296)

To be nutritious is to be usable in metabolism. The physicochemical properties of any given object, taken by themselves, do not render it intrinsically usable in metabolism. Nothing is non-relationally nutritious; things are only nutritious *for* organisms. Of course, being nutritious for any particular organism is nothing more than a matter of having the appropriate physical and chemical properties. But what makes those physicochemical properties appropriate is their metabolizability by that organism. Wood is nutritious for termites but not for humans, and peaches are nutritious for humans but not for termites. Analogously, to be colored is to be perceivable as colored. The physicochemical properties of any given object, taken by themselves, do not render it perceivable as colored. Things are only colored *for* organisms. Again, being colored any particular hue for any particular organism is nothing more than a matter of having appropriate physicochemical properties. But what makes those physicochemical properties appropriate is the fact that they allow that organism to perceive that color. Oranges are orange

for (most) humans, but they are not necessarily orange for other trichromats, much less for dichromats.

Nutritiousness and color are quintessentially *ecological* properties: there is no nutritious wood except in the environments of woodeaters like termites, and there are no colored objects except in the environments of color perceivers. Humans likely evolved trichromatic color vision partly to distinguish fruits from leaves (Mollon 1989; Jacobs 1996; Regan et al 2001). On this hypothesis, the perceivable colors raspberry red and leaf green evolved in tandem with the perceptual capacity of humans to discriminate between how the surfaces of raspberries and leaves respectively reflect light. In sum, color, realized by SSRs in relation to perceptual systems, is the ecological property that functions to enable organisms to visually discriminate environmental objects by hue.

While color relationalists tend to advocate relationalism as an alternative to physicalism about color, it would be perfectly coherent for them to embrace both, as accounts of distinct phenomena. A distinction can be drawn between the colors that organisms perceive and the intrinsic properties of objects that physicists might reasonably label 'colors'. Relationalism might be true of color qua perceivable property of objects—the 'color' phenomenon of interest in perceptual psychology—while physicalism is true of color qua intrinsic property of objects—the 'color' phenomenon of interest in the physics of light.

Relationalists need not take a stance on the issue of whether SSRs resemble (or correspond to) perceivable colors sufficiently to count as the intrinsic colors of physical objects. Instead, they can point to the double dissociation between SSRs and perceivable colors as reason to reject physicalism and embrace relationalism *as an account of perceivable color*. Perceivable colors are ecological properties relativized to particular perceivers and particular viewing conditions, whether or not SSRs (or their physico-chemical realizers) deserve to be labeled 'intrinsic colors' in their own right.

The rest of this article will draw from the argument for a relationalist view of color that I have just sketched in order to resist the conflation of attitudes of belief with intrinsic cognitive states of belief. Attitudes of belief are ecological properties, just like perceivable colors. Just as perceivable colors are realized by SSRs that function to make objects look colored to color perceivers, attitudes of belief are realized by dispositions to act, react, think, and feel that, taken together, function to render people *believers* in the eyes of belief attributors. And I will argue that, just as even committed physicalists about intrinsic colors ought also to accept relationalism about perceivable colors, even those philosophers who are committed to the existence of cognitive states of belief qua intrinsic features of organisms ought also to accept a relationalist account of attitudes of belief.[3]

---

[3] My argument for and account of attitudes of belief—while usefully illustrated by analogy with the argument for relationalism about color—does not stand or fall with any particular metaphysics of color. Like all analogies, the analogy between belief and color is imperfect and incomplete. Most strikingly, organisms *see* colors. Following Shannon Spaulding (2015), I reject views according to which belief attributors literally perceive beliefs. In some ways, this point of disanalogy renders the metaphysics of attitudes of belief easier to pin down than the metaphysics of perceivable colors. Embracing belief attribution as a thoroughly cognitive (as opposed to perceptual) phenomenon enables us to ignore tricky questions about the cognitive penetration of perception. It also enables model-theoretic accounts of belief

3. Belief

There is a problem for views that conflate attitudes of belief with intrinsic cognitive states of belief that parallels the problem for physicalism about perceivable color. The primary candidates for cognitive states of belief are brain states, subpersonal computational states, and personal functional states. But there are double dissociations between attitudes of belief—the beliefs that people attribute to each other in everyday life—and all three of these candidates for intrinsic cognitive states of belief, which mirror the double dissociation between perceivable colors and SSRs.[4] (Going forward, I will use Gilbert Ryle's label 'paramechanical' (1949) to pick out views which conflate attitudes with cognitive states.)

3.1. Against conflating attitudes with brain states

Paramechanical identity theorists type-identify attitudes of belief with brain processes, holding that beliefs "just *are* brain processes, not merely *correlated with* brain processes" (Smart 2007: 1).[5] Unfortunately for paramechanical identity theorists, there is a much-rehearsed double dissociation between attitudes of beliefs and the relevant neural processes and states of believers. First, just as perceived colors are multiply realizable by SSRs, attitudes of belief are multiply realizable by brain states (Putnam 1967). Second, on the least controversial of externalist assumptions, indistinguishable brain states can underlie distinct attitudes of belief when embedded in distinct environments (Putnam 1975).

The classic arguments for both multiple realizability and externalism hinge on the plausible claim that folk psychological belief attribution practices do not perfectly track neurophysiological similarities and differences. Insofar as these classic arguments hold water, the double dissociation arising therefrom indicates that *attitudes of belief* ought not be conflated with brain states. Insofar as the identity theory is the best metaphysics of *cognitive states of belief*—and a recent revival provides good reason to think it may be (Shapiro 2018; Thomson & Piccinini 2018)—we ought not conflate attitudes with cognitive states.

Paramechanists have responded to this double dissociation between attitudes of belief and brain states in several ways. Identity theorists downplay multiple realization and either deny

---

attribution (Godfrey-Smith 2005; Maibom 2009; Spaulding 2018; Curry forthcoming b), according to which interpreters construct and wield model psychological profiles of people in order to ascribe attitudes (and other traits) to those people. Model psychological profiles are more theoretically and empirically tractable than the amodal perceptual processes posited by theorists who countenance mindseeing. A related point of disanalogy stems from the impact that cultural forces have on models of belief. Culture *may* affect color perceptual learning (Connolly 2017), but models of belief are much more culturally variable (Heyes & Frith 2014; Lavelle 2019; Curry 2020). Likewise, belief attributors can construct models of new beliefs on the fly (Curry forthcoming a), whereas color perceivers cannot learn to perceive new colors.

[4] These dissociations have a cascading structure: although computationalism dodges the arguments against identity theory, and pure functionalism dodges the arguments against computationalism, the arguments against (paramechanical) pure functionalism also condemn (paramechanical) computationalism and the arguments against (paramechanical) computationalism also condemn (paramechanical) identity theory. Given this cascading structure, I will give more space to the later (more comprehensive) arguments.

[5] Unlike Smart, U.T. Place (1956) developed his pioneering version of the identity theory about consciousness as a supplement to his staunchly anti-paramechanical Rylean view of belief.

externalism or broaden the relevant physical states to include features of the environment. Dualists pull apart the physical and mental domains. Eliminativists deny that beliefs exist. However, by far the most popular way to respond to the double dissociation is to adopt paramechanical functionalism about belief. Paramechanical functionalism encompasses a diverse family of views which all hold that beliefs are individuated with respect to the functional roles they play in cognitive systems.

3.2. Against conflating attitudes with subpersonal computational states

Put overly simply, cognitive systems receive perceptual inputs and emit behavioral and experiential outputs. Computational functionalism is the doctrine that beliefs are functionally characterized subpersonal states: cogs that help cognitive machines transition from inputs to outputs.[6] According to one popular version of computational functionalism, beliefs are subpersonal relations to mental representations that play the psychofunctional role of disposing the believer to act as if the world were a particular way (Fodor 1987; Porot & Mandelbaum 2021).

Computational functionalists embrace the multiple realizability of beliefs by brain states. For the functionalist, brain states can multiply realize attitudes of belief because what makes a brain state realize a belief is a matter of cognitive function rather than physiology. Many computational functionalists are also happy to individuate cognitive states with respect to the external environment (Harman 1987; Block 1990; Kitcher 1991). Many computational functionalisms thus elegantly dodge the double dissociation between brain states and attitudes of belief, while reaffirming the conflation of attitudes of belief with cognitive states of belief. Nevertheless, paramechanical computational functionalists must reckon with a double dissociation of their own.

First, attitudes of belief are multiply realized by subpersonal computational states, just as computational states are multiply realized by brain states. As Hilary Putnam forcefully put the point, "there is absolutely no reason to believe that there is one computational state that all possible human beings who think that "there are lots of cats in the neighborhood" must be in" (1988: 104). Consider three people who all share the attitude of belief that a mug contains hot coffee. While her conscious mind is preoccupied with work, a distracted Delia orders a "dark roast for here" from Mo, the barista at her local café, and instinctively tells her three-year-old son, Roger, to "watch out; Mommy's coffee is very hot." Pouring the coffee, Mo the barista subpersonally computationally represents a 196°F French Roast made from Guatemalan beans being poured into a stoneware mug made by a local potter. Delia's relevant subpersonal representations are comparatively impoverished; she computes a dark roast coffee that she remembers liking the taste of, being poured into a nice mug, steaming enough that she should warn Roger. For his part, Roger, who has been raised alongside his mom's coffee habit, takes 'mug' to refer to any drinking vessel whatsoever, has no clue that there are different kinds of

---

[6] 'Computational functionalisms', as I use the term, include machine state functionalism, most classic forms of psychofunctionalism, and many embodied/embedded/extended/enactive/etc. functionalisms that complicate the simplistic input-computation-output functional analysis but retain some commitment to mental states intervening between perception and action. Computational functionalism is, however, to be distinguished from pure functionalism and teleofunctionalism, which I discuss in §§3.3 and 4.

coffee, and wields only a rough conception of degrees of heat.

Mo, Delia, and Roger all share the attitude of belief that there is hot coffee in the mug—that is, lay attributions of this belief to each of them would be veridical. But the functionally individuated computational states that underlie their respective beliefs diverge sharply, in terms of both representational content and psychofunctional role. Roger's notions of 'mug' and 'coffee' and 'hot' are different than Delia's, though not so different that it results in them having a different belief in this context.[7] Moreover, Roger's subpersonal computations over representations of the coffee and mug transform different inputs into different outputs than his mom's richer representations, and Mo's still richer representations transform still different inputs into different outputs. For example, Mo alone is prone to feel proud that the coffee is the optimal temperature for this particular Guatemalan French Roast, and Roger alone is prone to infer that every mug, cup, and glass in the café is a mug practically brimming over with scary hot coffee. The three café denizens' subpersonal computational states multiply realize their shared attitude of belief.[8]

The case for the multiple realization of attitudes of belief by subpersonal computational states is strengthened by cross-species comparisons. As Kim Sterelny (1990a; 1990b) argues, animals of different species often subpersonally realize (what can be fruitfully categorized as) the same attitudes via distinct computational states.

> Consider just the difference in perceptual structures between bats and owls. Owls have notoriously acute night vision, whereas bats find their way around by echo location. So if we had reason to attribute to bats and owls the same psychological state—say that they both perceive mice—then that state could hardly be individualistically defined. Perceptual systems vary greatly; their only common feature is that their function is the extraction of information for the adaptive control of behavior. (Sterelny 1990b: 98)

A bat and an owl can also share the attitude of believing there is a mouse in the grass, despite great variance in the computational states underlying the creatures' respective beliefs. There is no principled way of picking the bat or owl's (or forest ranger's) subpersonal relations to mental representations as *the* proper computational state to realize the belief that there is a mouse in the grass.

Second, attitudes of belief exhibit a *unity* (Curry forthcoming a) which may not exist among the computational states that subpersonally realize them. Consider my attitude of belief that I can help myself to La Colombe coffee from the Center for Neuroscience & Society lounge

---

[7] A thoroughgoing externalist might deny this difference in representational content (Burge 2010; Fodor & Pylyshyn 2016). I prefer externalisms which allow for internal factors that cause differences in content between mental representations with the same referent. Regardless, the claim that Mo, Delia, and Roger's mental representations have different content is not required to dissociate attitudes of belief from computational states: as explained in the next sentence of the main text, differences in psychofunctional roles—or in what Fodor calls "the syntactic structure of modes of presentation" (1992: 54), if you go in for that kind of thing—do the trick.

[8] Ryle (1949), Dennett (1978), Pylyshyn (1980; 1984), Putnam (1988), Schwitzgebel (2018), and myself (Curry forthcoming a) discuss similar cases.

on weekday mornings. This belief is concrete, specific, and coherent, but its cognitive underpinnings are complex. Believing it requires a cognitive system boasting a conjunct of functionally independent computational relations to mental representations, including representations of coffee and the company La Colombe and freeness of charge and permission and weekday mornings and my capacity to fetch things, as well a mental map of how to get to the relevant lounge.

There is unity in my attitude of belief—I live as if I can help myself to La Colombe from the Center for Neuroscience & Society lounge on weekday mornings in a unified pattern—but this unity plausibly emerges from disunity at the subpersonal level of computational states. Psychofunctional relations to my mental representation of weekday mornings are only peripherally associated with psychofunctional relations to my mental representation of freeness of charge. It is dubious that an explanatorily fecund cognitive architecture would (nonarbitrarily) tie these functionally independent computational states together as components of a single, functionally unified subpersonal cognitive state of belief. The only reason to tie them together would be that they both underlie an attitude of belief. But, on pain of begging the question of whether to conflate attitudes with cognitive states, that attitude of belief should not be conflated with a conjunction of computational states that are disunified at the subpersonal level. The paramechanist thus has no grounds for denying that the whole emergent pattern is more than the sum of its computational realizers.[9]

Indeed, the two prongs of the double dissociation between attitudes of belief and computational realizers are mutually reinforcing. If two different believers can share an emergent pattern despite differing in its computational realizers, then that pattern must be more than the sum of its parts. And if emergent patterns are what are of interest in everyday belief attribution, then there is no *prima facie* reason to doubt—and some additional reason to approbate—the extant evidence of multiple realization.

Especially in light of reasons to doubt that the folk construe beliefs as productive inner causes (Curry 2018), the multiple realization of attitudes of belief by (functionally disunified conjunctions of) subpersonal states should be enough to convince theorists to refrain from conflating attitudes with their computational realizers. And it has so convinced some paramechanists. Sterelny, for example, avoids the double dissociation between attitudes of belief and subpersonal computational states by instead conflating attitudes of belief with cognitive states of belief that play functional roles at the personal level of explanation. For high-level paramechanical functionalists like Sterelny, the bat, owl, and forest ranger can all be said to believe that there is a mouse in the grass insofar as they are disposed to behave, think, and feel in patterns that function to track the existence of the mouse in the grass. Paramechanists like Sterelny require that "the animals in question have certain discriminatory, memory or calculative abilities [to qualify as believers], but don't care [for the purpose of high-level belief attribution] how those abilities are computationally realized" (1990b: 99).

---

[9] Matthews (2007: 241) gives a very similar argument for the conclusion that attitudes must be personal (rather than subpersonal) phenomena, though (like Sterelny, as well as Jackson and Pettit, as discussed below) he goes in for a paramechanical view which conflates attitudes with personal functional states.

3.3. Against conflating attitudes with personal functional states

Frank Jackson and Philip Pettit call this a "purely functional theory", since it says that "to have beliefs and desires is to be understood purely functionally" (1990: 43): as having no relevant features over and above (or under and below) delivering certain outputs in response to certain inputs. Like Sterelny, Jackson and Pettit stress that the relevant functional roles mediating between inputs and outputs are *not* subpersonal psychofunctional roles of the sort that differentiate owls, bats, and forest rangers (1990: 34, 37). Instead, they are personal patterns of dispositions described at an abstract enough level to subsume the ways in which owls, bats, and forest rangers function to represent their environments. It does not matter whether disunified conjuncts of computational states multiply realize these abstractly characterized and purely functionally individuated beliefs; "it does not matter for the success of our passage back and forth between situations, behavior, beliefs, and desires how many states inside the agent are required to work the trick" (41). Insofar as agents do successfully work the trick—do go from inputs to outputs in patterns reliably identifiable as believing—those agents are believers.

On this (purportedly) purely functional person-level conception of cognitive states of belief, Mo, Delia, and Roger share the cognitive state of belief that there is hot coffee in the mug, and there is functional unity to my cognitive state of belief that I can help myself to La Colombe coffee from the neuroscience lounge on weekday mornings. Moreover, as Jackson and Pettit argue, it is difficult to deny that such personal functionally individuated beliefs exist.

The problem with Jackson and Pettit's view stems from the fact that, contra their rhetoric, there is no such thing as a *purely* functional analysis of agents, devoid of a principled criterion which determines which mediations between inputs and outputs are functionally relevant. In other words, the person-level functionalist must provide a criterion that fixes which of a believers' countless dispositions constitute any given belief; something has to explain why Delia's disposition to blow softly into her mug—but not her disposition to sneeze when placing her mug next to a vase of daisies—is partly constitutive of Delia functioning to represent the mug as containing hot coffee.

The diverse styles in which believers play person-level functional roles put this problem in high relief. The echolocating bat and sharp-eyed owl are disposed to transform different inputs into different outputs, yet their divergent dispositions functionally realize the same belief. Intraspecifically, Mo, Delia, and Roger possess different personal dispositions as well as different subpersonal computational states. Their divergent dispositions functionally realize the same belief, but they do not *purely* functionally realize the same belief. The aspiring person-level functionalist must give a criterion that non-arbitrarily lumps together Mo, Delia, and Roger's distinctly stylized patterns of living as fulfilling the same function for the respective believers.

Any of three candidate criteria would serve to explain what makes divergent patterns of dispositions all fulfill a particular person-level functional role, and thereby count as the same belief. First, divergent patterns of dispositions might stem from the same subpersonal computational state (or set of computational states). In that case, personal paramechanical functionalism collapses into subpersonal paramechanical functionalism, and faces the attendant double dissociation between attitudes of belief and cognitive states of belief. Second, divergent patterns of dispositions might serve the same irreducibly normative ecological purpose. Sterelny opts for this second, teleofunctional criterion. In §4, I will argue that teleofunctionalists have

further reason to refuse to conflate attitudes of belief with cognitive states of belief. Third, divergent patterns of dispositions might all sufficiently fit interpreters' models of belief or practices of belief attribution. Jackson and Pettit (belying the purported purity of their functionalism) opt for a paramechanical version of this third, commonsense criterion, arguing that "it is sufficient for having beliefs and desires that one be in states which satisfy the functional roles embodied in our everyday practice of predicting and explaining human behavior (for short, the folk roles)" (1990: 36).

Taking folk practices of belief attribution to supply the criteria that individuate beliefs does not make Jackson and Pettit interpretivists. On the contrary, Jackson and Pettit hold that "folk psychology is a *theory*" about purely functional person-level cognitive states that exist independently of practices of belief attribution (1990: 33). People have myriad cognitive states that nomically mediate between inputs and outputs. On Jackson and Pettit's view, folk psychology simply determines *which* independently existing cognitive states—which lawlike mediations between inputs and outputs—are captured by the commonsense theoretical positing of 'beliefs'. They embrace the paramechanical conclusion that attitudes of belief are just personal functionally individuated cognitive states that folk belief attributors designate 'beliefs'.

Unfortunately for paramechanical commonsense functionalists like Jackson and Pettit, there is a final double dissociation between attitudes of belief and the person-level functional states picked out by practices of belief attribution. Just as SSRs—surface-level functional states of objects—multiply realize perceivable colors in cases of metamerism, personal functional states multiply realize attitudes of belief. And just as the perceived colors of objects (but not SSRs) vary between perceivers, attitudes of belief (but not personal functional states) vary between attributors.

First, Jackson and Pettit's commonsense paramechanical functionalism fails to make the case that Mo, Delia, and Roger share a cognitive state of belief. The three café denizens have the same attitude of belief, as they are all in states which satisfy the folk role played by the attitude of believing there is coffee in the cup. According to Jackson and Pettitian commonsense functionalism, the three must therefore also share a cognitive state of belief constituted by whatever personal dispositions happen to mediate between inputs and outputs such that they function in the way theoretically posited by the folk role. But Mo, Delia, and Roger do not share the same cognitive state of belief; on the contrary, they share an attitude of belief *despite* having different stylized person-level functional configurations, just like metamers share a color despite boasting different SSRs.[10]

Attitudes of belief are individuated based on belief attributors' models or practices, whereas cognitive states of belief are individuated within the functional structure of believers. Jackson and Pettit (1990: 40–43) make a plausible (if not decisive) case that the predictive and explanatory success of attributions of attitudes of belief guarantees that people have cognitive states of belief—that ascriptions of attitudes of belief track some way or other in which believers are functionally configured. But there is no guarantee that all ascriptions of any given attitude track the same functional configuration. On the contrary, personal functional states multiply realize attitudes of belief. Mo, Delia, and Roger share an attitude despite having dramatically

---

[10] I provide more examples (and analysis) of diverse styles of belief elsewhere (Curry forthcoming a).

different dispositions. There is no principled way of picking Mo, Delia, or Roger's purely functional state as *the* proper personal functional state that is veridically captured by an ascription of attitude of belief.

Nor does the functional configuration of a believer fix which attitudes of belief she possesses. As I have argued at greater length elsewhere (Curry 2020), in some contexts it is intersubjectively indeterminate what somebody believes, just as the colors of The Dress are intersubjectively indeterminate. A single personal functional configuration sometimes underlies different attitudes of belief *for different belief attributors*. In such cases, the believer believes one thing for one belief attributor and another thing for another belief attributor, just as The Dress is white and gold for one color perceiver and blue and black for another color perceiver. There is no principled way of settling on one of these attitudes as the believer's one true belief.

Consider a variation on Daniel Dennett's (1998: 115) vignette about Ella, who has been behaving in troubling, self-undermining ways. Brown interprets Ella as believing that she ought to kill herself; Jones interprets Ella as believing, despite her angst, that she ought not kill herself. By stipulation, Brown and Jones have access to the full range of Ella's dispositions to act, react, think, and feel. Even so, they disagree. What Ella believes is intersubjectively indeterminate.

To flesh out the case, suppose Ella's most relevant dispositions are threefold: she feels no joy, it regularly occurs to her that she is capable of killing herself, she says she wants to die. Now suppose that Brown and Jones have developed different interpretive strategies—and attendant models of belief—over the course of their lives. Through his amateur study of human psychology and 19th century debates about natural selection, Jones has developed the deep conviction that people never believe they should kill themselves. Jones is convinced that evolutionary pressures have rendered people psychologically incapable of believing that they should end their own lives, though he allows that people sometimes mistake other beliefs—such as the belief that they would go to great lengths to escape their depression—for the belief that they ought to kill themselves. Jones models Ella as being confused about what she believes (and worries that she might act on the basis of that confusion) but does not model Ella as believing she ought to kill herself.[11] Brown, by contrast, interprets people as believing their assertions, absent probative evidence to the contrary, and holds that some people have genuine suicidal beliefs.

Jones voluminously and reliably predicts behaviors, thoughts, and feelings in accordance with his pet psychological theory (just look at how many people unsuccessfully attempt suicide, due to the lack of conviction with which they act!). Brown voluminously and reliably predicts

---

[11] Jones's may be a deeply flawed way of thinking about people's motivations. By stipulation, it is a predictively powerful interpretive strategy. (Jones predicts whether or not people will kill themselves with as much accuracy as Brown; he just does not use the attribution of belief that one ought to kill themselves in order to get to that prediction.) But perhaps it fails to capture what actually motivates people to act. However, whether Jones is a nonideal social cognizer is beside the point. Insofar as attitudes of belief are determined by folk models of belief—as Jackson and Pettit readily admit—they are determined by the nonideal, messy ways in which belief attributors actually model beliefs. Compare: humans may be deeply flawed color perceivers. We fail to represent whole chunks of the spectrum! But that humans are nonideal color perceivers is beside the point, when it comes to the metaphysics of perceivable colors. Perceivable colors are determined by the nonideal, messy visual spectrum, as it manifests itself in relation to particular nonideal color perceivers.

behaviors, thoughts, and feelings in accordance with his no-bullshit ethos. Ella believes she ought not kill herself from Jones's point of view, while believing that she ought to kill herself from Brown's point of view. Jones and Brown attribute different attitudes of belief despite agreeing on Ella's personal functional states.

What does Ella believe, *really*? Well, what color is The Dress, *really*? (If your kneejerk reaction is "it is really blue and black, just look at the dress in normal daylight viewing conditions!", what precise shades of blue and black is it? And is it really blue and black for bees as well as humans?) These questions are insufficiently precise. Ella believes different things—and believes things differently—for different belief attributors. The Dress is blue and black for 30% of the population and white and gold for 60% of the population. (The Dress is a special case, but even in normal daylight the dress is a different shade of blue for different color perceivers. And crocuses are different colors for different animals.) Everybody is right. The Dress *really* is white and gold for me, and it *really* is blue and black for people who see it as blue and black.[12] Likewise, Ella *really* believes she ought not kill herself for Jones, while *really* believing she ought to kill herself for Brown.

Complete knowledge of Ella's dispositions does not suffice to decide between these two interpretations, any more than knowledge of the dress's SSR suffices to decide its precise shade of perceivable blue. Ella has a single, intersubjectively stable set of dispositions to act, react, think, and feel, but she has different attitudes of belief for Brown and Jones, respectively. There is no principled way to settle on one of these attitudes of belief as *the* veridical representation of Ella's personal functional state, much less as Ella's one true belief.

As with color perception, there is interspecific as well as interpersonal indeterminacy in attitude attribution. Objects do not have the same perceivable colors for dichromats as for trichromats. Analogously, believers do not have the same attitudes for nonhuman attitude attributors as for humans. According to the currently mainstream interpretation of the empirical evidence on chimpanzee social cognitive capacities, chimps attribute goals, intentions, perception, knowledge, and ignorance to other animals, but do not attribute attitudes of belief (Call and Tomasello 2008). Even if recent reports that chimps do attribute attitudes of belief have merit (Crockford et al. 2012; Krupenye, Kano et al. 2016; Buttelmann et al. 2017), it is unlikely that they model beliefs in the same manner as humans. For example, chimps do not model beliefs as centrally involving the disposition to assent to linguistically structured propositions.

This difference in models of belief does not entail that chimps always fall short of veridically attributing the beliefs that believers *really* have. Instead, chimps (like humans) may be

---

[12] As previously noted in footnote 3, readers need not accept this relativism about perceivable colors in order to accept the analogous relativism about attitudes of belief. The analogy between belief and color is illuminating, but it does not do any argumentative work. Rather, the argument for relativism about belief formally mirrors the argument for relativism about color, but one argument may be sound and the other unsound due to divergent facts about the respective phenomena of interest. It may be that The Dress is *really* particular determinate shades of blue and black for all humans, even while attitudes of belief are fully intersubjectively indeterminate. Elsewhere (Curry 2020), I have argued at length against the ways in which other interpretivists have relativized attitudes of belief to intersubjectively determinate normative standards (like color relationalists who relativize colors to species-standard perceivers and viewing conditions).

excellent at attributing the attitudes it serves their species-specific purposes to attribute (Boesch & Boesch-Achermann 2000; Boesch 2009; Nagel 2011; Andrews 2018). There may be no interspecifically determinate answer to the question 'what attitudes does Chimpella *really* have?', just as there is not interspecifically determinate answer to the question 'what perceivable colors does the crocus *really* have?' Chimpella really has one set of attitudes of belief for Sally the human, and a different set for Red Peter the chimp. But Chimpella does not have an interspecifically variable personal functional state of belief. Thus, Chimpella's attitudes of belief ought not be conflated with her cognitive states of belief.

The reader need not accept this full-on relativism in order to see the problem intersubjective indeterminacy presents for conflating attitudes of belief with cognitive states of belief. An anti-relativist could, instead, treat intersubjective indeterminacy as grounds for irrealism rather than relativistic realism about attitudes of belief. Or else the anti-relativist might follow Dennett (1998) in claiming that what Ella believes is *objectively* (rather than merely *intersubjectively*) indeterminate. Finally, the anti-relativist might follow Donald Davidson (2001), Lynne Rudder Baker (1995), and Bruno Mölder (2010) in denying the coherence of the Ella and Chimpella cases—indeed, the coherence of intersubjective indeterminacy—on the grounds that veridical belief attribution is an irreducibly social, irreducibly normative, and uniquely human ability.[13] All of these theorists can agree that Jones and Brown (or Sally and Red Peter) attribute different attitudes of belief to (Chimp)Ella, and that there are no facts to be uncovered *about (Chimp)Ella's personal or subpersonal functional architecture* that fix one attribution as *the* veridical attribution.[14] Attitudes—if they exist—are metaphysically fixed in relation to interpretive capacities (or practices), not solely by personal functional roles, subpersonal computational states, or brain states. Moreover, brain states multiply realize computational states, which multiply realize pure functional roles, which multiply realize attitudes of belief. Thus, attitudes of belief ought not be conflated with intrinsic cognitive states of belief.

3.4. How attitudes of belief are like perceivable colors

My solution to the problems facing paramechanical theories of belief mimics the relationalist solution to the problem facing physicalist theories of color. Attitudes of belief, like

---

[13] For Davidson, it follows from the irreducibly social nature of belief attribution that if Jones and Brown know the same set of facts about Ella, then they cannot rationally disagree about what Ella believes. For Baker, Ella believes whatever the common sense interpretation would have her believe. Similarly, for Mölder, what Ella believes depends on a canonical ascription, which is partly determined by how ordinary people attribute beliefs. Elsewhere (Curry 2020), I have argued against these views and for relativism about belief. My present point is that attitudes ought not be conflated with cognitive states if attitudes are determined relative to attributors' models of belief, regardless of whether the veridicality conditions for belief attribution are set relativistically, objectively, or intersubjectively.

[14] For attitude irrealists, this is because no lay belief attributions are veridical. For Dennettians, this is because there is no objective fact of the matter. For Davidsonians, this is because what fixes the veridical attribution is a constitutive norm of interpretation—the principle of charity. For Bakerites, this is because what fixes the veridical attribution is what passes for common sense in a linguistic community, and, similarly, for Mölderians it is because what fixes the veridical attribution is how ordinary ascribers interpret people.

perceivable colors, are relational properties. Just as perceivable colors arise from the relationship between colored objects and color perceivers, attitudes of belief arise from the relationship between believers and belief attributors. In particular, believing is possessing a pattern of dispositions that a belief attributor recognizes as a taking of the world to be some way. Importantly, it is the *pattern* of dispositions—which emerges as an attitude of belief only in relation to a belief attributor's model—rather than the functionally individuated dispositions themselves, which belief attributors recognize as beliefs.

Thus stated, this minimal framework of a view is consistent with a wide range of dispositionalisms (Baker 1995; Schwitzgebel 2002) and interpretivisms (Dennett 1998; Davidson 2001; Mölder 2010) that carry debts to Ryle's (1949) landmark account of attitudes of belief.[15] But the analogy with color indicates a novel way of fleshing out a Rylean view. Attitudes of belief are ecological properties, like perceivable colors. Beliefs serve functional roles in the environments of mindreaders—organisms who attribute beliefs to themselves and others. In particular, attitudes of belief are the externalistically individuated patterns of dispositions in virtue of which believers—qua objects in social environments—function to seem to believe to belief attributors— qua subjects making sense of their social environments. Like nutritiousness and color (*mutatis mutandis*), attitudes of belief exist only in relation to organisms who have evolved to interact with their social environments by way of the attribution of belief. To believe that a mug contains hot coffee is to function, as an object in a social environment, to be attributed the belief that a mug contains hot coffee. Functioning in this way is a matter of having an appropriately interpretable pattern of dispositions to act, react, think, and feel: that is, a pattern of dispositions which the belief attributor would recognize as constitutive of belief.[16]

Paramechanists assume that attitudes of belief function as cogs in the cognitive systems of believers. This assumption is mistaken. Attitudes of belief do not contribute directly to the cognitive operations of the believers to whom they are attributed.[17] Cognitive states of belief (if they exist) do that productive causal work. Attitudes of belief, by contrast, function as characteristics of believers (including the belief attributor herself) in attributor-relative social environments.

Diverse styles of belief and intersubjective indeterminacy pose problems for paramechanists who take all beliefs to be intrinsic characteristics of believers, but not for interpretivists who refuse to conflate attitudes of belief with cognitive states of belief. Despite their disparate cognitive states, Mo, Delia, and Roger share the attitude of belief that there is coffee in the mug, insofar as they all live in ways that sufficiently fit the general-purpose model of that belief wielded by a belief attributor. And, as mentioned, interpretivists afford themselves the flexibility to account for—or dismiss—intersubjective indeterminacy in any of several ways.

---

[15] Elsewhere (Curry forthcoming b), I have argued that dispositionalists are *ipso facto* interpretivists, so I will henceforth restrict my discussion to interpretivism.

[16] In other publications, I have defended and fleshed out other aspects of this interpretivist account of belief (Curry 2018, 2020, forthcoming a, forthcoming b) and proposed an analogous interpretivist account of intelligence (Curry forthcoming c, forthcoming d).

[17] Attitudes of belief do, of course, contribute indirectly to cognitive operations. The bolt of self-knowledge that accompanies self-attributing the belief that one's job sucks might lead one to quit.

Dennett embraces (rare and practically insignificant) objective indeterminacy; Davidson insists that what somebody believes is objectively and intersubjectively determined by the principle of charity; Baker and Mölder advert to empirically informed common sense to resolve any apparent indeterminacy. My ecological interpretivism embraces the intersubjective indeterminacy of attitudes of belief while denying that there is any objective indeterminacy about what somebody believes. Ella objectively believes one thing for Jones and objectively believes another thing for Brown.

Interpretivists thus dodge the concerns about multiple realizability, externalism, and intersubjective indeterminacy that plague paramechanical versions of the identity theory, computational functionalism, and personal functionalism. But another popular version of the view that attitudes of belief are theoretically posited cogs in cognitive systems—paramechanical teleofunctionalism (Millikan 1984; Lycan 1987; Dretske 1988; Sterelny 1990b; Papineau 1993; Burge 2010; Neander 2017)—dodges these concerns almost as deftly. Happily, however, a critical discussion of how teleofunctionalism dodges these concerns will serve both to condemn its paramechanical incarnation and to add plausibility and detail to my ecological version of interpretivism.

4. Evolution

The central tenet of teleofunctionalism is that beliefs (and other content-bearing mental states) have irreducibly normative functions. According to paramechanical teleofunctionalists, beliefs *evolved*—and, in individual believers, develop—to be cogs in cognitive systems that serve representational and inferential purposes.

4.1 Paramechanical teleofunctionalism about belief

Whereas interpretivism is the theory that attitudes of belief are patterns of dispositions that belief attributors identify with taking the world to be some way, paramechanical teleofunctionalism is the theory that attitudes of belief are cognitive states that serve to make representations of believers' environments available for cognitive processing. The dominant, etiological form of teleofunctionalism dictates that beliefs have been naturally selected to serve the purposes played by believers' cognitive systems.

Teleofunctionalism differs from other functionalisms by placing an emphasis on the biological (as opposed to computational or pure dispositional) functions played by beliefs. As Elliot Sober quipped, teleofunctionalists put "the function back into functionalism" (1985: 175). Computational roles and dispositions are not usually described in normative terms; instead, the functionalists discussed in §3 make descriptive claims about how beliefs function to transform inputs into outputs. Teleofunctionalists, on the other hand, make irreducibly normative claims about how beliefs are supposed to function, given their evolutionary (and developmental) etiology. An attitude or cognitive state has a teleofunction insofar as it is constitutively aimed at fulfilling a biological purpose of an organism.

According to Ruth Millikan's influential etiological teleofunctionalism, beliefs are evolved cognitive means by which organisms assimilate information from their environments and figure out how to behave accordingly. Millikan writes that "the categories of intentional psychology are function categories in the biologist's sense of 'function', taking this to be a sense in which function

is determined by evolutionary history rather than by current dispositions" (1993: 171). What beliefs you have does not depend solely on the current computational (or personal functional) architecture of your cognitive system; it depends on the naturally teleological etiology of the mechanisms and cognitive states that system comprises.

> This means that categories such as belief … are biological-function categories—very broad and general ones, of course. Compare the categories limb, hormone, circulatory system, eye, visual system, etc. More contentious, the claim includes that such categories or types as belief-that-it-is-raining … are carved out with reference to biological functions. (1993: 172–173)[18]

In particular, for Millikan, "one of [belief]'s proper functions is to participate in inferences in such a manner as to help produce fulfillment of desires" (1993: 71). A believer has the belief-that-it-is-raining insofar as she has a cognitive state that was selected for the dual purposes of affirming the proposition that it is raining and enabling the believer to make useful inferences therefrom.

Paramechanical teleofunctionalism apparently avoids the problems with multiple realizability and intersubjective indeterminacy faced by paramechanical identity theorists, computational functionalists, and personal functionalists. Mo, Delia, and Roger share the same teleofunctional cognitive state of belief that the mug contains hot coffee, insofar as their respective computational (or dispositional) states all serve the same proper purpose. (Each of the café denizens' computational states also serves other purposes, teleofunctionally grounding other beliefs, but their respective cognitive states of belief that the mug contains hot coffee are all individuated by virtue of this shared purpose, abstracting away from implementational differences.) And paramechanical teleofunctionalists can take Ella to have a single, intersubjectively determinate attitude of belief by appealing neither to Ella's current cognitive architecture nor to intersubjective agreement, but to the evolutionary history of Ella's cognitive state. Jones and Brown both know how Ella actually functions, but the crucial question for the

---

[18] Millikan adds: "(though, in the case of beliefs, not directly according to function)" (1993: 173). This is because "being a little more precise, it is the belief-forming mechanisms that produce the adaptations, the adjustments, of the organism to the environment, the beliefs. Beliefs themselves are functionally classified, are "individuated," not directly by function but according to the special conditions corresponding to them that must be met in the world if it is to be possible for them to contribute to proper functioning of the larger system in a historically normal way" (189). For Millikan, belief-forming mechanisms and belief-consuming mechanisms are selected for teleofunctions, whereas beliefs are individuated on the basis of their ability—read: systemic capacity function (Cummins 1975)—to aid in the proper functioning of these cognitive mechanisms. This nuance of Millikan's teleofunctionalism does not impugn my analysis. For one thing, the idea of attitude-of-belief-forming- and attitude-of-belief-consuming-mechanisms is dubious precisely because we ought not conflate attitudes of belief with cognitive states of belief. Moreover, if my arguments are good, then insofar as it does make sense to speak of attitude-of-belief-forming-mechanisms, these mechanisms must be taken to include the social mindshaping forces detailed below. And insofar as it makes sense to speak of attitude-of-belief-consuming mechanisms, these mechanisms must be taken to include the belief attribution practices detailed below. Thus, if my arguments are good, then a follower of Millikan should take the mechanisms that form and consume cognitive states to have distinct proper functions from the mechanisms that form and consume attitudes.

teleofunctionalist is how she properly *ought* to function. If Jones's amateur evolutionary psychology is on the right track, then Ella's belief is anti-suicidal. Otherwise, Ella's attitude of belief is suicidal.[19] Either way, there is a single intersubjectively determinate fact about what Ella believes, as determined by the proper biological (rather than actual subpersonal or personal) function of Ella's cognitive state of belief.[20]

I previously emphasized dissociations between attitudes of belief and intrinsic cognitive states, analogous to the dissociations between perceivable colors and SSRs. In order to argue against paramechanical teleofunctionalism, I am going to exploit another dimension of the analogy between beliefs and colors. Just as colors co-evolved with organisms' capacities to see color, attitudes of belief co-evolved with organisms' capacities to attribute beliefs. Insofar as etiological teleofunctionalism is a theory of attitudes of belief as well as cognitive states of belief, teleofunctionalists ought to differentiate between the teleofunctions of the former and the teleofunctions of the latter. And insofar as it assigns distinct teleofunctions to attitudes of belief and cognitive states of belief, teleofunctionalism becomes a version of (rather than paramechanical rival to) my ecological interpretivism about attitudes of belief.

I am therefore not going to argue against teleofunctionalism, though I have reservations (especially about its etiological incarnations).[21] Instead, I am going to argue that committed teleofunctionalists should give up on their paramechanism: they should uphold the distinction between attitudes of belief and cognitive states of belief while being teleofunctionalists about both. My argument will hinge on questioning paramechanical assumptions about the functional roles attitudes of belief evolved to play, and whose purposes they evolved to serve. Attitudes of belief do not primarily serve believers; they did not evolve to be cogs in cognitive systems that properly function to help believers transform inputs into outputs and move around their environments. On the contrary, attitudes of belief evolved in tandem with belief attribution capacities, just as perceivable colors evolved in tandem with visual systems. *If* attitudes of belief have distinctive proper functions, then they properly function within the social environments of

---

[19] The notion that Ella's belief that she ought to kill herself plays a selected-for proper function is decidedly off-kilter. For interesting discussion of maladaptive (as well as selectively inert) beliefs, see Morgan (1883) and Peacocke (1992), as well as Millikan's (2000) response.

[20] Fodor (1992) has influentially argued that teleofunctionalism faces a different kind of indeterminacy worry, which he terms 'the disjunction problem': there is no way of saying whether a frog's belief is *there is a fly*, or *there is a bug*, or *there is a small black object*. Sterelny (1990: 125–127) provides an externalistic response to the disjunction problem, and Neander (2017) addresses it with reference to empirical findings about toad fly-detection capacities.

[21] My general reservations concerning etiological teleofunctionalism are articulated by Paul Sheldon Davies (2001, 2009), and my specific reservations concerning etiological teleofunctionalism about the individuation of belief are articulated by Sober (1985) and Sterelny (1990: 128–137). Note that, due to these reservations, the central argument of this section is conditional in form. *If* one is a teleofunctionalist, then they ought to be an interpretivist about attitudes of belief too. If one is not a teleofunctionalist, then the arguments of the previous sections ought to have already convinced them of interpretivism about attitudes of belief. This section should, nevertheless, be of interest to all readers, insofar as exploring the evolutionary history and present functions of everyday belief attribution adds plausibility and detail to my ecological interpretivism (regardless of whether one buys teleofunctionalism as a principle for individuating beliefs).

belief attributors, as means by which attributors regulate (McGeer 2007), manipulate (Malle 2004), predict (Dennett 1998), explain (Andrews 2012), and (ethically, aesthetically, epistemically, and pragmatically) evaluate whole patterns of thoughts, feelings, and behaviors.

4.2. Against conflating attitudes with teleofunctional cognitive states

Humans evolved not only to interact with inanimate objects in their environments, but also to interact with other minded organisms. In Tad Zawidzki's (2013: 233) words, "the most important features of most primate ecologies are social" in a sophisticated sense: they involve primates' recognition of others as end-driven agents. Human social cognition comprises an especially sophisticated set of capacities that were made possible because humans evolved in especially rich social environments.[22] Zawidzki calls this set of capacities the "human sociocognitive syndrome," and takes it to include pervasive cooperation, language use, sophisticated mindshaping ability, and sophisticated mindreading ability.[23]

Zawidzki calls mindreading "sophisticated" when it involves the explicit attribution of attitudes, as opposed to merely "*some kind of appreciation* of conspecifics' [attitudes] … a kind of sensitivity to or ability to track at least some propositional attitudes" (2013: 13).[24] To be able to mindread, in this sophisticated sense, is to be able to attribute attitudes of belief veridically. Though we do not know the precise evolutionary and cultural forces that have led to the belief attribution abilities of modern humans, we have good reason to believe that the capacity arose out of the selective pressures of our historical social environments. When you interact with people a lot, it is immensely useful to be able to figure out what they believe. Knowing that a dominant peer does not believe the food you crave exists can help you acquire much needed extra nutrients (Hare et al. 2000). For mothers, reliably attributing beliefs to other adults is crucial to figuring out who to trust to help rear vulnerable, slow-developing human babies (Hrdy 2009). More generally, being able to figure out what weird things our conspecifics believe allows us to offer satisfactory explanations of their anomalous behavior (Andrews 2012: 224–230). Similarly, attributing beliefs to ourselves—and telling our friends what we believe—helps us portray our own behavior as rational and responsible (Malle, Knobe & Nelson 2007). Zawidzki argues that "our ancestors first started attributing full-blown propositional attitudes … to rehabilitate status in the wake of apparently counternormative behavior, especially apparent reneging on explicit commitments"

---

[22] Other proponents of the importance of social environments in human evolution who have been influential on my thinking include Tomasello (1999), Hrdy (2009), Apperly (2011), Andrews (2012), Sterelny (2012), Chase (2013) and Heyes (2018).

[23] Many theorists countenance these four capacities as (at least nigh) uniquely human. There is, however, no consensus about which of these capacities emerged first in the history of human evolution. Many psychologists and philosophers take mindreading to have led to the other abilities, but Zawidzki argues that mindshaping is "our sociocognitive linchpin" (2013: 1). Sterelny (2012) and Heyes (2018) rally against the idea of any single magic bullet.

[24] Although he stresses that the point of sophisticated mindreading is not to identify cognitive states (2013: 237), Zawidzki nevertheless fails to question the paramechanical assumption that attitudes of belief do happen to be "concrete, unobservable causes of behavior" (2013: 11). This leads him to endorse a Dennettian interpretivism about the targets of unsophisticated mindreading while retaining certain paramechanistic assumptions about the attitudes of belief attributed in sophisticated mindreading.

(2013: 224). All in all, the ability to attribute beliefs was plainly adaptive in our ancestral social environments.

Being mindread*able* was adaptive too. If you think, feel, act, and react in patterns that belief attributors can latch onto, then you are more likely to be trusted with precious food, babies, and promises. Developing understandable (and nonthreatening) patterns of dispositions is crucial to being admitted into human societies in which your conspecifics have your back. Crucially, though, these attitudes of belief help believers out only insofar as they render them more intelligible in the eyes of belief attributors (including the believer herself).

These mutually reinforcing adaptive benefits—of the capacity to attribute beliefs and the propensity to live in patterns understandable as beliefs—led to an evolutionary ratchet effect. Sophisticated 'mindshaping' (Mameli 2001) comprises a set of practices including imitation, pedagogy, conformity to norms, and narrative self-constitution. What these practices have in common is that they function to make patterns of behavior more homogenous across a population. In other words, mindshaping delimits the patterns of dispositions to act, react, think, and feel that normal members of a community are likely to have. On Zawidzki's account, sophisticated mindshaping made the evolution of sophisticated mindreading possible. In order to veridically attribute beliefs to other members of their social environments, humans had to first use mindshaping techniques to construct those environments such that beliefs would manifest themselves as understandable—if sometimes aberrant—patterns of behaviors, thoughts, and feelings. The capacity to attribute beliefs in turn helped people regulate, manipulate, predict, explain, and evaluate the particular thoughts, feelings, and behaviors of their peers, on the basis of these patterns, which allowed for ever more sophisticated imitation, pedagogy, conformity, and narrative self-constitution practices. Through this more sophisticated mindshaping, humans developed more refined, more normalized, and thus more predictable, attitudes of belief. Attitudes of belief and mindreading abilities therefore (both biologically and culturally) co-evolved.

Zawidzki's theory of the evolution of social cognition suggests that attitudes of belief were naturally selected in virtue of the functional roles they play in mindreaders' social environments. It was adaptive for our ancestors to be attributed beliefs (and judged favorably on account of those beliefs), just as it was adaptive to attribute beliefs veridically. In Millikan's terms, attitudes of belief have the proper function of enabling belief attribution, which in turn has the proper function of helping belief attributors predict, explain, regulate, and evaluate the thoughts, feelings, and behaviors of believers. To believe that it is raining is to have a pattern of dispositions demarcated by the teleofunction of enabling sophisticated mindreaders to attribute the belief that it is raining to you.

The natural selection of stable attitudes of belief—qua patterns of dispositions to act, react, think, and feel—inevitably involved the alteration of cognitive states of believers. Dispositions to act, react, think and feel are, after all, ontologically dependent on the cognitive systems that enable organisms to act, react, think, and feel. Nevertheless, it is not cognitive states of believers that were selected for when mindshaping practices made it increasingly adaptive to behave, think, and feel similarly to everybody else. Successful mindreaders pick up on attitudes of belief qua patterns that fit their models of belief, not qua cognitive states underlying those patterns. It is thus *mindreadable* (rather than purely functional) patterns of dispositions that fulfill the

teleofunction of enabling mindreading.

Consider the analogy with the evolution of color perception. As mentioned, there is good evidence that primate trichromacy co-evolved with the colors of the fruits that primates eat. It was adaptive for trichromatic primates to locate nutritious fruits, and it was adaptive for colorful-fruit-bearing plants to disperse their seeds via primate digestive systems. This co-evolution inevitably involved the alteration of the intrinsic physicochemical properties and SSRs of fruits. But those physicochemical properties and SSRs were not selected for. The properties of the fruits that were selected for were their perceivable colors, because perceivable colors are the properties that serve the teleofunction of enabling color perceivers to locate fruits (and thereby disperse their seeds).

The moral of this evolutionary fable is that, because colors exist for the purpose of being seen, they cannot be reduced to *invisible* physicochemical properties of colored objects. Indeed, it is not any non-relational property of fruits, but the way fruits *look* red *to primates* that conferred a selective advantage. Perceivable colors are for color perceivers—they exist in order to be seen by creatures like us—and it is only through being for perceivers that they help plants disperse their seeds.

Analogously, on an account that blends interpretivism and teleofunctionalism about belief, attitudes of belief are patterns of dispositions to act, react, think, and feel that were selected for the teleofunction of making people recognizable as believers to belief attributors. There are no perceivable colors without (at least historical) color perceivers, and there are no attitudes of belief without (at least historical) belief attributors. Attitudes of belief are for belief attributors—they exist in order to be grasped by creatures like us—and it is only through being for attributors that they help believers navigate their social environments.

There might be cognitive states of belief with the teleofunctions assigned to them by philosophers like Millikan. But if such cognitive states of belief exist, they plausibly acquired their etiological teleofunctions long before belief attribution capacities and attitudes of belief co-evolved. Humans (or their ancestors) had need of cognitive states serving representational and inferential purposes, to help them interact with their environments, long before they developed sophisticated social cognition. The two varieties of belief have distinct proper functions, and thus deserve distinct teleofunctional analyses. Whereas cognitive states of belief fulfill (or fail to fulfill) their teleofunctions within the proprietary cognitive systems of believers, attitudes of belief fulfill (or fail to fulfill) their teleofunctions within the social environments of belief attributors. Committed teleofunctionalists should therefore reject paramechanism in favor of a teleofunctional version of interpretivism about attitudes of belief (while, if they like, retaining their classic teleofunctionalism about cognitive states of belief).[25]

Cognitive states of belief may hew closely to attitudes of belief. Indeed, the evolution of

---

[25] It might be objected that attitudes of belief function to give attributors a grip on the cognitive functioning of believers. Attitudes of belief *can* function, *in part*, to give attributors a loose, indirect grip on the cognitive functioning of believers. But I reject the implication that attitudes of belief can therefore unproblematically be conflated with cognitive states of belief. After all, perceivable colors can function, in part, to give perceivers a grip on how objects reflect light. (Seeing that my mug is blue gives me a loose, indirect grip on the fact that it reflects more light at the lower end of the visual spectrum.) Nevertheless, perceivable colors cannot be unproblematically conflated with SSRs.

mindreadable attitudes of belief may have (contingently) depended on attitudes of belief coming to weakly supervene on cognitive states of belief, such that, given the course of cognitive evolutionary history, believers with the same cognitive states could not have different attitudes of belief. Fred Dretske has argued that beliefs weakly supervene on brain states via an analogy with how monetary value weakly supervenes on the physical properties of coins and bills. The sociohistorical etiology of money guarantees the latter supervenience relation—"thanks to the government's efforts, every piece of paper that has a particular set of intrinsic properties is a genuine $20 bill" (2000: 269). Dretske proposes that the phylogenetic and ontogenetic etiology of beliefs analogously guarantees the contingent but complete supervenience of the mind on the brain. One could run a very similar argument for the weak supervenience of attitudes of belief on cognitive states of belief.[26] Especially if cognitive states of belief were pitched at the personal level (á la Pettit and Jackson 1990) and non-etiologically teleofunctionally individuated (á la Sterelny 1990a), this argument might be persuasive. I thus will not argue that attitudes of belief cannot be given a reductive analysis in terms of cognitive states of belief. Instead, I have argued that attitudes ought not be conflated with cognitive states. But reducibility should not be assumed either, given how little is currently known about cognitive architecture, as well as the even murkier question of how attributor-relative aspects of mind are coordinated with intrinsic aspects of mind.

Whether attitudes of belief weakly supervene on cognitive states of belief is an open empirical question. At the other extreme, eliminativists have long stressed that it is an open empirical question whether cognitive states of belief exist at all. If it turns out that there is nothing to be found in human cognitive systems resembling (much less supervened upon by) attitudes of belief, then we should be eliminativists about cognitive states of belief. Nevertheless, the attitudes of belief that play functional roles in the social environments of belief attributors will remain ontologically unscathed. If it turns out that attitudes of belief do weakly supervene on cognitive states of belief, then functionalists will be vindicated in their realism about cognitive states of belief. But they will not be vindicated in their paramechanism. Weak supervenience does not amount to type-identity (Haugeland 1982). Functionalists type-identify beliefs via functions, and attitudes of belief play different functions (in different systems) than cognitive states of belief. Socially primitive creatures living in a world without belief attributors could have evolved cognitive states of belief, but it would have been impossible for them to evolve attitudes of belief.

Indeed, assigning distinct teleofunctions to attitudes and cognitive states restores the double dissociation between the two. Cognitive states of belief that have evolved to serve representational and inferential teleofunctions plausibly multiply realize attitudes of belief that serve teleofunctions of rendering people recognizable as believers. Moreover, believers with intersubjectively determinate cognitive states of belief can have intersubjectively indeterminate attitudes of belief, since they can have distinct attitudes with teleofunctions serving the distinct purposes of different belief attributors.

---

[26] Dretske makes a distinction between behavior and bodily movement and argues that beliefs cause behavior whereas the brain states they weakly supervene on cause bodily movement. Depending on how it goes, an argument for the weak supervenience of attitudes of belief on cognitive states of belief might entail that whereas cognitive states of belief cause behavior, attitudes of belief render that behavior suitable for folk explanation, prediction, regulation, and evaluation.

Readers who are uneasy with just-so evolutionary explanations of psychological traits need not be alarmed by the adaptationist bent of the preceding discussion. We can see who color and belief are for without reference to evolutionary history. Right now, somewhere in the Blue Ridge Mountains, a berry is being eaten by a bear—leading to the dispersal of its seeds—because of its vibrant color. The functional role of the color is to signal EAT ME! to the bear. Analogously, when we attribute the belief that the mug contains hot coffee to Mo, Delia, and Roger, we are not speculating about the trio's respective cognitive architectures. We are making sense of, and coming to terms with, their complex patterns of thoughts, feelings, actions, and reactions. The functional role of these patterns is to signal I TAKE THE MUG TO CONTAIN HOT COFFEE! to themselves and other belief attributors. Likewise, the question of what Ella believes matters to Brown and Jones: they are trying to figure out how to intervene, on the basis of whether or not she believes she should kill herself. They are not speculating about what productively causes her to act—by stipulation, they agree about that. Instead, they are arguing about how best to characterize her in terms of belief, so that they can lend a helping hand more effectively.

Teleofunctionalists ought not conflate attitudes of belief with cognitive states of belief; the former emerge within belief attribution practices, whereas the latter emerge within (etiological or systemic) analyses of cognitive systems. By refusing to conflate attitudes with cognitive states, philosophers can provide a rigorous account of the former that is invulnerable to revision at the whim of ephemeral trends in theories of cognitive architecture. Identity theorists, computational functionalists, commonsense functionalists, teleofunctionalists, and eliminativists about cognitive states of belief should all agree that attitudes of belief are patterns of living that play functional roles in belief attributors' social environments. Pace recent paramechanical critiques (Crane 2016; Quilty-Dunn & Mandelbaum 2018), interpretivism complements the cognitive scientific study of belief, just as relationalism complements the physical study of color.

5. Interpretivism all the way down?

I have argued that attitudes of belief emerge in relation to interpretation, and must therefore be distinguished from any intrinsic cognitive states of belief posited by cognitive scientists. I have not yet discussed the more radical possibility that cognitive states of belief themselves are not intrinsic characteristics of believers—that is, that cognitive states of belief, like attitudes of belief, emerge in relation to interpretation. Perhaps cognitive states of belief emerge relative to philosophers and scientists' theoretically motivated belief attribution practices, in much the same manner as attitudes of belief emerge relative to folks' socially motivated belief attribution practices. Philosophers of various persuasions might find this claim plausible due to their specific commitments about the nature of belief, or due to a more general commitment to interpretivism about the mental (á la Mölder 2010) or to perspectivism about functional individuation (á la Cummins 1983 or Craver 2007) or to perspectivism about all scientific kinds (á la Kuhn 1962 or Giere 2010).[27]

The possibility that cognitive states of belief are themselves interpretation-dependent does not necessarily obviate the distinction between attitudes of belief and cognitive states of

---

[27] Lee and Dewhurst (2021) offer an illuminating discussion of the relationship between interpretivism (spelled out in terms of Dennett's 'intentional stance') and perspectivisms about mechanistic explanation.

belief. After all, the theoretical motivations that drive scientists' belief attribution practices might be sufficiently different from the social motivations that drive folks' belief attribution practices to make it such that the cognitive states of belief that emerge in relation to the former are importantly distinct from the attitudes of belief that emerge in relation to the latter. As mentioned at the outset of this article, there is good reason to doubt that folks ordinarily construe beliefs as inner causes that produce behavior. But that is precisely how many cognitive scientists construe beliefs. Thus, even if they are interpretation-dependent, cognitive states are plausibly individuated with respect to the functional role that they purportedly play within a cognitive system (which itself may be individuated in relation to cognitive scientists' perspectives and explanatory aims)—and can be empirically discovered to exist, or not, via investigations into whether the cognitive system functions in the theorized manner. In contrast, as I argued in §4, attitudes of belief are individuated with respect to the role they play (not in the believer's cognitive system but) in the social environment of the belief attributor. Given this core difference in how cognitive states of belief and attitudes of belief are individuated—which holds even if both varieties of belief emerge in relation to interpretation—the distinction between them strikes me as iron-clad.

Nevertheless, I will not pretend to have settled the issue of whether it would make practical sense to continue to distinguish cognitive states of belief from attitudes of belief if one were to accept the more radical claim that all believing is interpretation-dependent. Instead, I will be content noting that I do take myself to have shown that the distinction between cognitive states of belief and attitudes of belief is extremely useful within our present intellectual environment in which many philosophers and scientists remain committed to the existence of non-interpretation-dependent beliefs as central theoretical posits in cognitive science. Determining whether cognitive states of belief—if they exist—are intrinsic or interpretation-dependent is a tall order. It implicates old, vexed debates about the mind-body problem, the relation of mind to world, natural kinds, natural teleology, and the theory-ladenness of science, as well as newer vexed debates about cognitive architecture, mechanistic explanation, and the explanatory scope of cognitive science. Those debates have not been settled in this article. If they are eventually settled in a manner that supports both the existence and the interpretation-dependence of cognitive states of belief, then it is possible that my distinction between attitudes and cognitive states will have been obviated (my remarks in the previous paragraph notwithstanding).

As those debates currently stand, however, many philosophers remain convinced that cognitive states of belief are intrinsic characteristics of believers (perhaps with some important characteristics derived from their etiology). Jerry Fodor once wrote that if something like this were not "literally true, then practically everything I believe about anything is false and it's the end of the world" (Fodor 1989: 77). In this article, I have argued that if Fodor (or Jackson and Pettit, or Millikan, or any other philosopher who posits non-interpretation-dependent cognitive states of belief) is correct about cognitive states of belief, then they ought not conflate them with attitudes of belief. Rather, all philosophers should be ecological interpretivists about attitudes of belief, no matter their views about the beliefs invoked in cognitive science, just as all philosophers should be relationalists about perceivable color, no matter their views about the colors invoked in the physics of light. Even if it does not hold fast in the end, the working distinction between attitudes of belief and cognitive states of belief is, at present, indispensable as a means of guiding everybody to the truth about the beliefs we human beings attribute to each other in everyday life.

References

Almagro Holgado, M. & Fernandez Castro, V. (2020). The Social Cover View: a Non-Epistemic Approach to Mindreading. *Philosophia* 48: 483–505.

Andrews, K. (2012). *Do Apes Read Minds?* MIT Press.

Andrews, K. (2018). Apes track false beliefs but might not understand them. *Learning & Behavior*, 46: 1, 3–4.

Baker, L. R. (1995). *Explaining Attitudes*. MIT Press.

Block, N. (1990). Inverted Earth. In J. Tomberlin (Ed.), *Philosophical Perspectives*. Ridgeview Press.

Boesch, C. and Boesch-Achermann, H. (2000). *The Chimpanzees of the Taï Forest: Behavioural Ecology and Evolution*. OUP.

Boesch, C. (2009). *Wild Cultures*. CUP.

Boghossian, P. & Velleman, D. (1989). Colour as a secondary quality. *Mind* 98:81-103.

Burge, T. (2010). *Origins of Objectivity*. OUP.

Buttelmann, D., Buttelmann, F., Carpenter, M., Call, J., & Tomasello, M. (2017). "Great apes distinguish true from false beliefs in an interactive helping task". *PLoS ONE* 12(4): e0173793.

Call, J. & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Science*, 12(5):187–192.

Chase, P. (2013). Human Culture is more than Memes in Transmission. In G. Hatfield and H. Pittman (Eds.), *Evolution of Mind, Brain and Culture*, Penn, 347–378.

Chirimuuta, M. (2015). *Outside Color*. MIT Press.

Churchland, P. (1981). Eliminative Materialism and the Propositional Attitudes. *Journal of*

Philosophy Vol. 78, No. 2., 67–90.

Cohen, J. (2009). *The Red and the Real.* OUP.

Connolly, K. (2019). *Perceptual Learning*. OUP.

Crane, T. (2016). *The Mechanical Mind*, Third Edition. Routledge.

Craver, C. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. OUP.

Crockford, C., Wittig, R., Mundry, R., & K. Zuberbuhler (2012). Wild Chimpanzees Inform Ignorant Group Members of Danger. *Current Biology* 22:2, 142–146.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy* 72:741–60.

Cummins, R. (1983). *The Nature of Psychological Explanation*. MIT Press.

Curry, D. S. (2018). Beliefs as inner causes: the (lack of) evidence. *Philosophical Psychology* 31:6, 850–877.

Curry, D. S. (2020). Interpretivism and norms. *Philosophical Studies* 177:4, 905–930.

Curry, D. S. (forthcoming a). Belief in character studies. *American Philosophical Quarterly*.

Curry, D. S. (forthcoming b). Interpretivism without judgement-dependence. *Philosophia*.

Curry, D. S. (forthcoming c). Street smarts. *Synthese*.

Curry, D. S. (forthcoming d). *g* as bridge model. *Philosophy of Science*.

Davidson, D. (1963). "Actions, Reasons, and Causes". *Journal of Philosophy*, 60(23), 685–700.

Davidson, D. (2001). *Subjective, Intersubjective, Objective*. OUP.

Davies, P. S. (2001). *Norms of Nature: Naturalism and the Nature of Functions.* MIT Press.

Davies, P. S. (2009). *Subjects of the World.* Chicago University Press.

Dennett, D. C. (1978). *Brainstorms: Philosophical Essays on Mind and Psychology*. MIT Press.

Dennett, D. C. (1998). *Brainchildren*. MIT Press.

Dewhurst, J. (2017). Folk psychology and the Bayesian brain. In Metzinger & Wiese (Eds.), *Philosophy and Predictive Processing*.

Dretske, F. (1988). *Explaining Behavior.* MIT Press.

Dretske, F. (2000). *Perception, Knowledge and Belief.* Cambridge Studies in Philosophy.

Fodor, J. (1987). *Psychosemantics.* MIT Press.

Fodor, J. (1989). Making Mind Matter More. *Philosophical Topics* 17: 1, 59–79.

Fodor, J. (1992). *The Elm and the Expert*. MIT Press

Fodor, J. & Pylyshyn Z. (2016). *Minds without Meanings*. MIT Press.

Gert, J. (2017). *Primitive Colors*. OUP.

Giere, R. (2010). *Scientific Perspectivism*. Chicago.

Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature*. CUP.

Godfrey-Smith, P. (2005). Folk Psychology as a Model. *Philosophers' Imprint* 5:6, 1–16.

Hare, B., Call, J., Agnetta, B. & Tomasello, M. (2000). Chimpanees know what conspecifics do and do not see. *Animal Behaviour* 59 (4): 771–785.

Hardin, C. L. (1988). *Color for Philosophers: Unweaving the Rainbow*. Hackett.

Harman, G. (1987). (Nonsolipsistic) Conceptual Role Semantics. In Ernie LePore (Ed.), *New Directions in Semantics* (pp. 55–81). London: Academic Press.

Hatfield, G. (1992). Color Perception and Neural Encoding: Does Metameric Matching Entail a Loss of Information? *PSA 1992: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, David Hull & Mickey Forbes (eds.), 1:492–504.

Hatfield, G. (2009). *Perception & Cognition*. OUP.

Haugeland, J. (1982). Weak supervenience. *American Philosophical Quarterly*, 19: 93–103.

Heyes, C. & Frith, C. (2014). The cultural evolution of mind reading. *Nature*, Vol 344, 6190.

Heyes, C. (2018). *Cognitive Gadgets: The Cultural Evolution of Thinking*. HUP.

Hilbert, D. (1987). *Color and Color Perception: A Study in Anthropocentric Realism.* CSLI.

Hrdy, S. (2009). *Mothers and Others*. Harvard University Press

Hutto, D. (2011). Presumptuous naturalism: a cautionary tale. *American Philosophical Quarterly*, 48(2): 129-145.

Jackson, F. (1977). *Perception: A Representative Theory*. Cambridge University Press.

Jackson, F. & Pettit, P. (1990). In defence of folk psychology. *Philosophical Studies* 59: 31–54.

Jacobs, G. H. (1996). Primate photopigments and primate color vision. *Proceedings of the National Academy of Science USA* 93:577–81.

Kitcher, P. (1991). Narrow Psychology and Wide Functionalism. In R. Boyd, P. Gasper, and J.D. Trout, (Eds.), *The Philosophy of Science* (671–685). MIT Press.

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago.

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, 354(6308), 110–114.

Lavelle, J. S. (2019). The impact of culture on mindreading. *Synthese*. Online First.

Lee, J. & Dewhurst, J. (2021). The mechanistic stance. *European Journal for Philosophy of Science* 11 (1):1–21.

Lycan, W. (1987). *Consciousness*. MIT Press.

Maibom, H. (2009). In Defence of (Model) Theory Theory. *Journal of Consciousness Studies*, 16, 360.

Malle, B. (2004). *How the Mind Explains Behavior*. MIT Press.

Malle, B., Knobe, J., & Nelson, S. (2007). Actor-observer asymmetries in behavior explanations: New answers to an old question. *Journal of Personality and Social Psychology*, *93*, 491–514.

Mameli, M. (2001). Mindreading, mindshaping, and evolution. *Biology and Philosophy,* 16, 597–628.

Matthews, R. (2007). *The Measure of Mind*. OUP.

McGeer, V. (2007). The regulative dimension of folk psychology. In D. Hutto, M. M. Ratcliffe, (Eds.), *Folk Psychology Re Assessed*, Springer, 137–156.

Menzel, R.; Steinmann, E.; Souza, J. De; Backhaus, W. (1988). Spectral Sensitivity of Photoreceptors and Colour Vision in the Solitary Bee, Osmia Rufa. *Journal of Experimental Biology* 136 (1): 35–52.

Millikan, R. (1984). *Language, Thought and Other Biological Categories*. MIT Press.

Millikan, R. (1993). *White Queen Psychology and Other Essays for Alice*. MIT Press

Millikan, R. (2000). *On Clear and Confused Ideas: An Essay about Substance Concepts*. Cambridge University Press.

Mölder, B. (2010). *Mind Ascribed*. John Benjamins Publishing.

Mollon, J. D. (1989). "Tho' she kneel'd in that place where they grew..." The uses and origins of primate colour vision. *Journal of Experimental Biology* 146:21–38.

Morgan, C. L. (1883). Suicide of Scorpions. *Nature* 27, 313–314.

Nagel, J. (2011). Knowledge as a mental state. *Oxford Studies in Epistemology* Vol 4, Ch 10.

Neander, K. (2017). *A Mark of the Mental: In Defense of Informational Teleosemantics*. MIT Press.

Papineau, D. (1993). *Philosophical Naturalism*. Blackwell.

Peacocke, C. (1992). *A Study of Concepts*. MIT Press.

Place, U. T. (1956) Is consciousness a brain process? *British Journal of Psychology* 47:44–50.

Porot, N. & Mandelbaum, E. (2021). The science of belief: A progress report. *WIREs Cognitive Science* 12(2): e1539.

Poslajko, K. (2020). The Lycan-Stich argument and the plasticity of "belief". *Erkenntnis*. Online First.

Putnam, H. (1967). Psychological predicates. In Capitan & Merrill (eds.) *Art, Mind, and Religion* (37–48). Pittsburgh.

Putnam, H. (1975). The Meaning of meaning. *Minnesota Studies in the Philosophy of Science* 7:131–193.

Putnam, H. (1988). *Representation and Reality*. MIT Press.

Pylyshyn, Z. (1980). Cognition and Computation. *Behavioral and Brain Sciences* 3, 111–32.

Pylyshyn, Z. (1984). *Computation and Cognition*. MIT Press.

Quilty-Dunn, J. & Mandelbaum, E. (2018). Against dispositionalism: belief in cognitive science. *Philosophical Studies* 175: 2353–2372.

Rabin, J., Houser, B., Talbert, C., & Patel, R. (2016). Blue-Black or White-Gold? Early Stage Processing and the Color of 'The Dress'. *PLoS ONE* 11(8).

Regan, B.C, C. Julliot, B. Simmen, F. Viénot, P. Charles–Dominique, J. D. Mollon (2001). Fruits, foliage and the evolution of primate colour vision. P*hil. Trans. R. Soc. Lond. B* 356: 229–283.

Ryle, G. (1949). *The Concept of Mind*. Chicago University Press.

Schlaffke, L., Golisch, A., Haag, L., Lenz, M., Heba, S., Lissek, S., Schmidt-Wilcke, T., Eysel, U., Tegenthoff, M. (2015). The brain's dress code: How The Dress allows to decode the neuronal pathway of an optical illusion. *Cortex*, 73: 271–275.

Schwitzgebel, E. (2002). A Phenomenal, Dispositional Account of Belief. *Nous*, 36, 249–275.

Schwitzgebel, E. (2013). A Dispositional Approach to Attitudes: Thinking Outside the Belief Box. In N. Nottelmann (Ed.), *New Essays on Belief* (75–99). Palgrave Macmillan.

Schwitzgebel, E. (2018) It's not just one thing, to believe there's a gas station on the corner. *The Splintered Mind* weblog, February 28. Retrieved from http://schwitzsplinters.blogspot.com/2018/02/its-not-just-one-thing-to-believe.html.

Shapiro, Lawrence (2018). Reduction redux. *Studies in History and Philosophy of Science Part A,* Volume 68, 10–19.

Smart, J. C. C. (1961). Colours. *Philosophy* 36:128–42.

Smart, J. C. C. (2000). The Mind/Brain Identity Theory. *The Stanford Encyclopedia of Philosophy.*

Sober, E. (1985). Panglossian Functionalism and the Philosophy of Mind. *Synthese,* 64, 165–94.

Sommers, F. (2009). Dissonant beliefs. *Analysis* 69:2, 267–274

Spaulding, S. (2015). On Direct Social Perception. *Consciousness and Cognition* 36: 472–482.

Spaulding, S. (2018). *How We Understand Others*. Routledge.

Sterelny, K. (1990a). Animals and Individualism. *Vancouver Studies in CogSci* 1, 323–339.

Sterelny, K. (1990b). *The Representational Theory of Mind*. Blackwell.

Sterelny, K. (2012). *The Evolved Apprentice.* MIT Press.

Strand, M. & Lizardo, O. (2015). Beyond world images: Belief as embodied action in the world. *Sociological Theory* 33:44-70.

Tanney, J. (2013). *Rules, Reason, and Self-Knowledge.* HUP.

Thompson, E. (1995). *Color Vision.* Routledge.

Thomson, E. & Piccinini, G. (2018). Neural representations observed. *Minds and Machines* 28 (1): 191–235.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition.* HUP.

Wallisch, P. (2017). Illumination assumptions account for individual differences in the perceptual interpretation of a profoundly ambiguous stimulus in the color domain: 'The Dress. *Journal of Vision* 17:5 1–14.

Webster, M. (2015). Individual Differences in Color Vision. *Handbook of Color Psychology*, A.J. Elliot, M.D. Fairchild, and A. Franklin (Eds.), Cambridge University Press, 197–215.

Yablo, S. (1995). Singling out properties. *Philosophical Perspectives*, 9, 477–502.

Zawidzki, T. (2013). *Mindshaping*. MIT Press.