



Faces and Situational Agency



uncorrected draft - please cite published version!!

Faces and Situational Agency

M. Crippen and G. Rolla

Abstract

Though there are many challenges to Ekman's thesis that there are basic emotions with universal corresponding facial expressions, our main criticism revolves around the extent to which grounding situations alter how people read faces. To that end, we recruit testifying experimental studies that show identical faces expressing varying emotions when contextualized differently. Rather than dismissing these as illusions, we start with the position—generally favored by embodied thinkers—that situations are primary: they are where specifiable and hence knowable properties first show up. We further argue that situationally inflected emotional expressions are informationally meaningful. We reject the idea that reading expressions is primarily about ascertaining internal mental states, arguing instead that people are registering overall situations when looking at faces. However, if mind is understood as a situated phenomenon that extends into active ecological frames, then one can still argue that mindreading is going on. Although we do not claim isolated things like cliffs or cars have agency, we speculate networked systems with cliffs, people, cars, bears, etc., collectively function with intentionality, more so if advancing a robust situated mind thesis, contra figures like Dennett who argue that people over-impute mind to things. Our position has practical implications insofar as it casts doubts on recent attempts to develop AI systems that extract emotional intent out of facial expressions since many of these systems are grounded on Ekman's basic view.

Introduction

Scholars have highlighted problems with the thesis that there are basic emotions with universal corresponding facial expressions, prominently defended by Paul Ekman (e.g., 1972, 1992, 2016) and marketed as a map for extracting hidden feelings from outward behavior. One issue is that commentators have not completely settled on what emotions are (Berridge 2018). Another is that standard lists of six-to-twelve fundamental, discrete emotions seems artificial and contrived (see Feldman Barrett et al. 2009). Yet another is that some expressions vary with culture (e.g., Jack et al. 2012; Crivelli et al. 2016)—not that defenders of the basic view completely ignore this last point. Still another objection is that investigative methods identifying basic emotions use frozen snapshots of exaggerated expressions almost never encountered in actual life, also funneling responses into limited numbers of options that bias results towards identifying small numbers of basic emotions (e.g., Russell 1993, 1994, 2006; Crippen 2021).

There are accordingly numerous reasons to object to Ekman's thesis, but our main criticisms are focused. We start by granting that faces convey lots of information, yet stress that this occurs in grounding situations that alter how expressions appear. Scenarios like this pervade everyday life where

changed context can make the same conventionalized expression look friendly or sexually aggressive, pain-stricken or ecstatic (Heaven 2020; Crippen 2021), and legitimately so, as we will argue. Our account has practical implications insofar as it casts doubts on AI systems designed to read emotional expressions that are almost exclusively grounded in Ekman-like theories. Though worth discussing in its own right, we emphasize how the shortcomings of AI here serve to highlight problems in basic emotion-expression research agendas, simultaneously promoting our claim that situations inflect how faces show up.

Our position is not just that situations change what we see on faces, but that such cases are not merely psychological projections, that is, consequences of how we represent things in our heads. After detailing how context changes how faces show up, we defend the position—usually favored by embodied thinkers—that situations are primary: they are where determinable and hence knowable properties first manifest. Whether talking about Gestalt scenarios, color constancy or an object’s length varying depending on velocity relative to the thing observed, many specifiable qualities emerge only within interrelational frames. This is similar to what Gibson’s ecological psychology (1966, 1979) calls affordances, defined as the possibilities of action that an environment allows a creature to perform and therefore realities that exist in relation to particular organisms. Situationally inflected emotional expressions are similar not only in that they require an observer, but also because they arise out of interrelations within the environment. Importantly, situationally inflected emotional expressions are also informationally meaningful: a hungry smile conveys different information depending on whether a middle-aged man directs it at a buffet table or a 14-year-old teenager in a bikini. This line of thinking augments research that grasps expressions as “socially shared conceptual knowledge that perceivers use to create meaning from ... physical changes” (Feldman Barrett 2012, p. 413). In a way this claim parallels Wittgenstein’s (1953) observation that grammar and concepts are necessarily shared and public, and also gets near to Gibson’s (1966, 1979) assertion that meaning is an ecological

phenomenon as opposed to being in the heads of agents, which is our argument about situationally inflected expressions.

The views we advance on how facial expressions are manifested are consistent with a situated understanding of mind, that is, the claim that the mind is embodied and extends into surrounding sociocultural and physical structures, which scaffold and constitute psychic functions. Building on this line, and the fact that context changes how emotions show up on faces, we reject the idea that reading expressions is primarily about interpreting internal mental states. We argue instead that people are registering overall situations when looking at faces—that they are sensing whether action is warranted, and if so, what kind. At the same time, if mind is understood as a situated affair that extends into active ecological frames, then one can still argue mindreading is occurring. Although we do not go so far as to claim cliffs have agency, it remains obvious that natural geographies, forests, people, bears and the like can be hostile or non-threatening. Moreover, taken as networked and dynamic ecologies that form parts of minds, an environment and things in it together function as a collection of interconnected cognitive systems. This is not only illustrated through human examples, but also reinforced by behavior-based AI, all of this challenging figures like Dennett (2006) who argue that people over-impute mentality to things.

Situating Faces

Since the 1960s, many ranking psychologists have maintained that people experience basic emotions, conventionally listed as six: anger, disgust, fear, happiness, sadness and surprise, but with variations in the exact number. A second part of the thesis is each basic emotion has a universal corresponding facial expression (e.g., Tomkins and McCarter 1964; Ekman 1972, 1992, 2016; Buck 1994; Smith et al. 2005; Jack et al. 2016). “Basic” means cleanly categorizable and hence discrete; the term additionally designates key emotions as fundamental and sometimes as building blocks for more complex ones (e.g., Du et al. 2014), comparable to primary colors like red and blue mixing to form

purple (Gu et al. 2019), all of this in a British empiricist vein. The outlook is mechanistic to the extent that it asserts that basic expressions occur involuntarily, proponents proposing detectable differences between genuine and feigned emotional facial behavior. Despite enjoying longstanding dominance, the basic emotions-expressions view has had challengers for decades. One line of criticism is that settings, body language, manners of interacting and hence the temporal unfolding of situations have all been shown to change how facial expressions show up (Aviezer et al. 2012a b; Heaven 2020; Crippen 2021). For this objection to be damning, however, contextual inflection needs to mark a meaningful change in the emotional situation, for example, a shift from safeness to riskiness. We aim to demonstrate this, but before we do so, we will highlight ways in which situations change the emotional expressions we see.

For about a century, filmmakers have recognized that context inflects the appearance of faces. Perhaps the best-known case is the Kuleshov effect, a cinematic phenomenon wherein an impassive expression in a single shot shows up with different emotions depending on how the scene is cut and thus how the face is situated (see Pudovkin 1926/1968, p. 168). The same thing happens to expressive faces, so that a shot of a man smiling may look friendly or dirty depending on whether it is cut to a mother with a child or bikini-clad teenager (see Crippen 2021).¹ A similar example is an experiment with participants reporting Darth Vader's mask going from neutral to sarcastic after some creative re-editing (Crippen 2021).²

Though people in Kuleshov-like scenarios do not command their changing facial appearances between different cuts, this does not affirm Ekman's thesis that expressions for basic emotions are largely involuntary and his corollary claim that observers can detect false emotional exhibitions. The obvious reason is that alterations undergone have nothing to do with muscles automatically jerking—even if this sometimes occurs in everyday life—since the same shots of performers are used.

¹ For bikini example, visit <https://www.youtube.com/watch?v=hCAE0t6KwJY>.

² For Vader example, visit <https://www.youtube.com/watch?v=kRvvIRbHfqM>.

Alterations instead follow from context inflecting expressions. Cases like this, moreover, are not just a cinematic anomaly, but have everyday validity. Heaven (2020) has compiled comparable real-life cases. One of them is a basketball player celebrating a slam dunk, whose expression looks pain-stricken when isolated from the context. Similarly, there are photographs of Adele receiving an award, Mexican spectators enjoying a football victory and Justin Bieber fans overcome by his presence, all of whom look sad if cut off from the broader circumstances. Indeed, if one was told that the Mexican team had lost, then most would see despairing expressions, meaning narrational context can affect interpretations of facial behavior (see Carroll and Russell 1996; Crippen 2021). The expressiveness of intonation also changes with the scenario, so that a dog's yelps may be joyful or mournful depending on whether a beloved master has returned or died (Crippen 2021). Situationally inflected expressions can therefore be added to other well-known Gestalt examples showing 1) that perception is multistable, that is, prone to appear in different ways; 2) that perception entails wholes that change the nature of individual parts; and 3) that context constrains possibilities of multistability (see Koffka 1935, Ch. 4; Köhler 1947, Chs. 5-6).

Ranking scholars have responded (e.g., Carroll 1993; Ekman and Friesen 2003) that emotional expression is only somewhat underdetermined by faces, also downplaying context (Ekman and O'Sullivan 1988). We grant that multistability has limits: a smile will not morph into a frown, and while Necker cubes—a commonly given example—may flip so that different planes appear forward-facing, they never manifest as pizzas. Yet studies do show that joyful people look pain- or grief-stricken when cut off from environmental contexts (Heaven 2020); or that deflated body language morphs orgasmic and victorious expressions into agony or defeat (Aviezer et al. 2012a b).³ An Ekmanian theorist could nonetheless respond that a contextual shift from joyful happiness to vicious or

³ Similarly, Peter Hujar's *Orgasmic Man* series consists of photographs of faces of men having orgasms, some of whom look sad or agonized precisely when background contexts are missing. One of these portraits was used for the cover of Hanya Yanagihara's novel, *A Little Life*.

lustful happiness still has a basic emotion at its foundation, i.e., happiness. However, the response assumes that one can offer a previous (*a priori*) distinction between basic and non-basic expressions, and that some expressions are ambiguous, and others are not. But the very point raised by cases we have cited is that these distinctions are arbitrarily drawn, with emotions like joking happiness, sarcasm, snide disgust, aggressive jubilation and angry glee overlapping (Crippen 2021). More problematically, the Ekmanian rebuttal does not address how situations can reveal joy when people otherwise falsely perceive grief in decontextualized faces, which amounts to a shift between two purportedly basic emotions. Neurobiological research also weighs against modular or discrete and hence basic emotions (Russell 2006; Feldman Barrett et al. 2009). Nor is there compelling evidence of universal expressions for epistemically nuanced emotions like shame or sarcastic disdain, which are relevant to deception detection. Unsurprisingly, then, recent reviews suggest Ekman's techniques for emotion and deception detection do not work (Weinberger 2010; Barrett Fieldman 2019; Heaven 2020).

Despite all these objections, Ekman's outlook remains popular. The general adherence to his view is in fact amplified by recent attempts to develop AI systems that extract emotional intent out of facial expressions since many of these systems are neural nets taught on training sets that presume Ekman's basic view (see Crawford 2021, Ch. 5). One reason is that even if Ekman's conveniently simple model were true, it would already be computationally demanding to read emotions from expression. Creating a system that can handle cases in which situations inflect expressions in informationally meaningful ways is all the more difficult. Again, in everyday life, parents may legitimately perceive differences between an old man smiling longingly at the ocean vs. their 14-year-old daughter, just as tears at weddings and funerals mean different things. A further obstacle to AI is that conditions in which context-dependent expressiveness occur can be unique, as with Vader's sarcasm, which is again informationally meaningful insofar as it reflects a maddening situation. Consequently, it can be difficult to gather sets of similar instances of such cases for training neural networks, which are currently favored for extracting emotions from facial behavior. In the existential

terms of AI critics like Dreyfus (2007), the human ability to rapidly negotiate facial expressions is a kind of practical know-how. Put otherwise, being able to immediately grasp onetime situations like the one with Vader is a skill that gets closer to art than formal science. As such, it is difficult to specify a general method for interpreting certain situations, even though people grasp them easily.

Contextual Meanings and Realities

So far we have shown that situational inflection occurs and it is informationally meaningful, as in the case of the man smiling hungrily at the teenager vs. the ocean. We want to push the thesis further, and argue that expressive changes arising through situational inflection are comparable to affordances. Such an outlook would grant, on the one hand, that situational inflection depends on the presence of an observer. On the other hand, the position would locate situational inflection in ecological relations and not exclusively in people's heads. This has metaphysical and epistemological ramifications that we introduce in this section, developing them further in subsequent ones, with the ultimate aim of explicating expressiveness within situated mind theory.

Gibson (1979, p. 129) asserts that affordances bridge “the dichotomy of subjective-objective.” In other words, affordances are behavioral facts about how subjects can act in environments and hence have a subjective side; yet they nonetheless “are in a sense objective, real” because they are really in the world but are co-determined by an organism's capacities, so that water is objectively walkable for some insects, yet not for humans (Gibson and Pick 2000, Ch. 2). Gibson regards affordance detection as discernment of ambient information, for instance, light by which a creature sees a path as traversable. Emotions affect the salience but not the presence of affordances on his view, so that a love of hiking makes a passage up a hill stand out without impacting whether or not it is there. Simultaneously, emotions in a way alter affordances by changing energy levels and hence the actions a setting avails (see Riener et al. 2011), for instance, making a rugged passage objectively less walkable and hence less passage-like to a depressed individual lacking energy. Gibson (1966, p. 285) equates

affordances to “values,” defined as “what things furnish, for good or ill.” His last book (1979) elaborates: “all these benefits and injuries, these safeties and dangers, these positive and negative affordances are properties of things *taken with reference to an observer* but are not properties of *the experiences of the observer*. They are not subjective values” (p. 137). This outlook “implies that the ‘values’ and ‘meanings’ of things in the environment can be directly perceived” (p 127). The idea, in other words, is that values and meanings exist, like affordances, in the coupling between agent and environment, a point we aim to extend to emotional qualities that show up in situations. We do this with the goal of building to the idea that what we often regard as mindreading is less about inferring hidden facts from faces and more about practically negotiating overt realities in situations, which themselves can be regarded as cognitive systems.

We begin our argumentative line with a defensible assumption: that perception is contextual. An example of a sodium-vapor lamp illuminating white denim illustrates. In this case, the cloth does not show up as orange-yellow, even though this light frequency strikes the retina. People instead almost always see it as white fabric bathed in orange-yellow, that is, unless an experimenter has them observing an isolated patch through a narrow aperture. This phenomenon, known as color constancy, tracks to the actual pigment of objects regardless of ambient light conditions, so that seeing the denim as white means registering what is not received from frequencies reflected from isolated cloth patches. Situationally inflected expressions are similar: we accurately register emotions by perceiving what is absent from isolated faces, as demonstrated in the earlier examples of elated people looking mournful when cut off from contexts, all this fitting Gestalt descriptions.

In the late 19th and early 20th centuries, one finds figures from many quarters detailing scenarios in Gestalt veins. Peirce (1878) asserts that we understand “hardness” not as an isolated property, but as the relational fact that one object is not easily scratched by another. Dewey (1929) notes likewise that scientists cause particles to appear by colliding them. He adds this is not peculiar because we regularly come to register things better by shaking or thumping them, or otherwise altering

our relation to them, as when magnifying starlight. Collingwood's (1939) question-and-answer manner of textual interpretation highlights that conceptual meaning, too, is relational and indeed multistable, with Wittgenstein (1953) doing similarly. This is such that the import of the sentence, *The ring is in the garbage*, varies depending on whether it answers the question, *Where's your wedding ring?* or *Where's that toy Green Lantern ring you found on the sidewalk?* (Crippen 2019). Merleau-Ponty (1945/1962, pp. 367-368) details perceptual variants of the above scenario, observing that the smoothness of wood is not a collection of separate tactile pressures; it is instead the way that a surface coordinates the hand into a style of engagement. This style, which cannot be broken into discrete parts, brings forth smooth textures, but also might engender sinewy toughness if scratched with a fishhook (Crippen 2017).

Commentators sometimes argue that multistability is not an ecological fact about the world in Gibson's sense, but instead a matter of how the brain organizes "inner" representations (e.g., Stadler and Kruse 1995). Our purpose is not to completely deny the existence of inner representations, yet we would point out that the multiple realizations of wood are a matter of actions in the world and thus explicable in terms of Gibson's ecological psychology. Our challenge is to explain how situationally inflected facial expressions are likewise external. Dewey (1929) offers one line of argument by referencing non-classical physics in defence of his situational philosophy. To consider one case, relativity and spectral shift describe instances in which length and color vary with relative velocity. This means that entities only have a specifiable length or color within a frame of reference, so properties show up differently depending on whether they are observed from a stationary standpoint or a high velocity relative to the object. Contextual inflection of faces is of course not entirely akin to the sort that physicists ponder. While both are relational, the multistability of expression requires a human perceiver, plus changed relations between a face and context, whereas the aforesaid measurement of light and spatial dimensions only presupposes a point of reference, though it is hard to see how either can be specified without an observer. In any event, it is arbitrary to conclude that something ceases to be real merely because agents are introduced into the relationships that allow it to emerge. Thompson

(2004, p. 386), one of the originators of enactive cognitive science, highlights this when asserting that the nutritional properties of sugar are not intrinsic to its molecular structure, but have “significance or value as food ... only in the milieu that the organism itself brings into existence.” The take-home lesson is that the continuous exchanges between inhabitants and environments literally bring forth and concretely generate realities that are externally extant, palpable, publicly available, valuable and hence meaningful to participants in the relationships (Rolla and Figueiredo 2021).

The notion of “physiognomic qualities,” introduced by Werner (1927) and elaborated by Gestalt psychologists such as Koffka (1935, Chs. 8, 9, 14) helps solidify why we think situationally inflected expressions are ecological realities. The concept does not refer to expressions built up out of separate elements like a crooked nose and furrowed brow. The term instead designates expressive wholes that are not reducible to the sum of their parts. Critically, the concept applies to both environments and expressive behavior, and characterizes the fact that we can perceive a gloomy forest or face when happy (Koffka 1935, Ch. 8), meaning we detect emotional qualities outside of us (Dewey 1934, Ch. 4; Dreyfus 1991, Ch. 10). Further, the idea of physiognomic qualities is not in the vein of contour or resemblance theories from aesthetics (e.g., Davies 2006), which assert we perceive certain cloud formations as angry because they mirror perturbed faces. Instead and without denying that contour cases sometimes hold, the Gestalt assertion is that overall configurations exhibit certain mood traits. Such happens when a cheerful pub district and the jovial faces of a group of revelers take on threatening tones when night falls, but more for a woman who objectively faces greater risk than her male counterpart. Here, the woman—who we will call Salma—finds herself in an unsettling public mood, as opposed to discovering a mood is inside her (see Dewey 1934, Ch. 4; Dreyfus 1991, Ch. 10). The “mood” can be grasped as an atmosphere. To the extent that the situation warrants vigilance from Salma, the revellers’ contextually inflected countenances add to the atmosphere regardless of whether any specific individual is threatening. The atmosphere therefore depends on the embodied conditions that affect behavioral dispositions that are constrained and influenced by objective aspects of the

environmental milieu, accordingly bridging the subjective-objective dichotomy in both enactive and ecological senses. The enactive side is that the atmosphere is concretely generated by Salma's embodied disposition, in the same way a depressed individual genuinely confronts an emotionally strenuous world because the affliction makes acting harder. The atmosphere is simultaneously enacted by the way others probe her with their gaze in the particular sociophysical setting, which inflects but also is modulated by their faces and behavior. The ecological side is that the atmosphere approximates affordances and has roughly the same ontological status as them. Gibson (1966, 1979) in fact advances a compatible view insofar as he repeatedly frames dangers as negative affordances because they are bad for some agent and thus make a setting less usable for that agent.

The foregoing introduces argumentative lines for claiming that environmental emotional characters and contextually inflected faces are subspecies of the same phenomenon 1) because both are relational gestalts and 2) because perception of either orients around registering possible actions. Take as an example the emotional quality of mysteriousness, which can equally manifest in a trail winding out of view into a forested hill or in the visage of an elegantly clad individual with braided purple hair leaning against a polished wood bar as jazz saturates the air. Here, the emotional character emerges from total arrangements, as opposed to a face, path, forest or hill alone. "Mysteriousness" designates an emotional quality we observe in settings or other people and not in us (Rosar 1994), and, in the case of the path, emerges partly as an outcome of the trail offering new discoveries around the next bend, while simultaneously assuring that explorers will not get lost (Kaplan and Kaplan 1989; Crippen and Schulkin 2020, Ch. 3). Something similar holds with the captivating individual, whose purple pleats in the context of a free jazz performance introduce intriguing wildness, yet with the elegant clothing and venue offering assurance that an encounter in the bar will not get completely out of hand.

Mysteriousness therefore manifests according to what we might do in a setting and with people and things in it, particularly when opportunities for approach and exploration are pronounced. In other words, the mysteriousness of the face ties to the intrigue of the overall setting, and can be grasped in

terms of affordances and values. As an affordance, mysteriousness is, in the phraseology of Gibson's (1979, p. 129) ecological psychology, equally about the environment and behavior. At the same time, and while we recognize that some enactivists and ecological psychologists butt heads (we think wrongly, but outside our current scope), the scenario is also one in which agents bring forth realities. Dewey (1934, p. 177) expresses the idea in his proto-enactive work, writing that "perception and its object are built up and completed in one and the same continuing operation." As such, experienced "qualities never were 'in' the organism," but are instead "qualities of interactions in which both extra-organic things and organisms partake" (Dewey 1925, p. 259).

Situated Minds

One point we have stressed is that the world is emotionally infused, at least for humans and many other cephalic creatures. In itself, this is not controversial, but the typical view is that we perceive, for example, sadness in things like weeping willows as a byproduct of social perception. Although this sometimes occurs, we have suggested the reverse is simultaneously so: that our ability to read expressive meaning on faces may be part of a more general ability to sense emotional attributes of settings. Seen thus, predatory looks may be like hostile settings, just as mysterious smiles and scenes may be alike because both promise something worthwhile while not revealing too much. None of this denies that faces convey a lot of information. However, except in laboratory experiments, faces are not encountered in isolation, and their expressiveness almost invariably relates to situations. We said at the outset that scholars like Dennett (2006) suggest that people over-impute mind to things. We propose an alternative based in situated mind theory: that environments are parts of cognitive or mental systems, "cognitive" or "mental" here used loosely to connote psychic functions. Such would supply explanation for cross-cultural beliefs in animism and the tendency to see expressive qualities in non-human things. To that end, we trace out situated mind theory in this section, connecting it to contextual

inflection of facial expressions. In the next section we frame mindreading as a sensing-out of situations that are themselves parts of extended cognitive systems.

In addition to earlier-introduced enactive standpoints, situated theories include embodied, embedded and extended varieties, collectively known as 4E cognitive science (Rowlands 2010). Embodiment stresses that psychic life follows from neural and extra-neural bodily structures engaging with the world (Chemero 2009). Embeddedness highlights the role that the environment plays in scaffolding mental processes (Wilson 2004). Enactivism adds that actions bring forth properties and experiences (Varela et al. 1991). The extended view, in its turn, asserts that mental operations extend into surroundings; for example, framing notepads as external memory enhancers or canes as vehicles through which blind people engender spatial perception (Clark and Chalmers 1998). Though often mutually implying, proponents of these various perspectives debate certain things like the importance or irrelevance of brain-based representations (Hutto and Myin 2013). We do not wish to rehearse the disagreements, but instead draw from these different positions when useful, partly on the premise that psychic life is pluralistic and such that it cannot be explained by any single outlook.

At the same time, we extract a key idea advanced unexceptionally by 4E advocates: that bodily structure and objects encountered limit and enable the way we can engage with things, bringing non-random structure to doings and undergoings and hence to mental encounters. A standard illustration has already been given: the hand caressing an object to realize smoothness, which is neither in the thing alone nor a mere brain-based representation. Another example is sensory substitution devices where head-mounted cameras stimulate skin or tongue. People actively exploring surroundings acquire an analogue of vision (see Varela et al., 1991, Ch. 8), so that body-environment interactions determine the perceptual mode and the properties manifested. Body-world interactions can be scaled up, with monarch butterflies migrating via light and magnetic cues, plus mountain ranges funneling them towards their destination (Crippen 2016). In fact, even single-celled creatures behave intelligently (e.g., Dewey 1920, Ch. 4; Di Paolo et al. 2017, Ch. 5). *Physarum polycephalum* is noted for using

slime trails and tubules. These mechanisms allow groups of these brainless organisms to externally mark where they have already travelled and also map food sources and even solve laboratory mazes (Nakagaki et al. 2000), in effect enacting affordances. By mechanisms like this, *P. polycephalum* can successfully navigate mazes, which would be cognitively taxing for humans without an aerial view (Crippen 2020, 2021). Faces—and we will expand on this later—can similarly be regarded as external scaffolds in that they coordinate group behavior (see Dumouchel 2022), supplying a punctuated, emotional map of surroundings, indicating safeties, dangers and other values.

Before their explicit formulation in the late 20th century, situated views had traction among behavior-oriented roboticists, who have acted on the premise that intelligence is not restricted to central processing units or CPUs. During the 1950s, researchers built devices that navigated environments via detection-actuation mechanisms (see Walter 1950; Boden 2006, Ch. 4). A decade later, Simon's (1969, Ch. 3) landmark treatment of artificial intelligence offered an oft-quoted example where the complexity of an insect's movement follows from the complicated contours of the terrain over which it crawls. A similar view was formulated a year earlier in a 1968 technical paper from Stanford University (Feigenbaum 1968) that noted it is easier to extract information from the environment than construct a world model inside a robot AI. This means “the most economic and efficient store of information about the real world is the real world itself” (p. 10), a claim that Brooks (e.g., 1990, p. 5), a Stanford grad and the foremost situated AI advocate, echoes verbatim. The contributions of Brook's research team include the robot vacuum, Roomba. Certain models have simple programming layers such as one causing it to rotate partly towards an obstacle that it hits and then move forward. Combined with the vacuum's round body and detection-actuation mechanisms, the device can thereby follow the straight line of a wall, the circular shape of a pillar or the right angle of a corner. The vacuum accordingly uses the room and things in it as a navigation grid, much as *P. polycephalum* employs slime and tubules. The robot's circular shape—which facilitates rotation and

thereby allows the vacuum to follow various dimensions—is critical to its minimally intelligent actions.

The point we are building to and will expand in the next section is that cognition, perception and emotion—in short, mind—complete themselves in the environment. The environment, in turn, forms a dynamic loop that responds to and interjects action-eliciting pressures on the agent, as in the earlier example of the woman negotiating the probing faces of revelers. The environment thereby functions as part of an extended intentional system, which raises the possibility that it can be mindread. This is not to say that we specifically endorse Clark and Chalmer’s (1998) extended mind hypothesis, for its functionalist assumptions have historically been too closely tied to cognitivism and representationalism for our tastes. An approach that better fits our ideas is the scaffolded mind hypothesis (see Sterelny 2010), according to which our cognitive capacities depend on and can be transformed by the environment. This stance is encompassing enough to accommodate the fact that there are times when the external environment is literally a part of the cognitive system. The earlier-discussed *P. polycephalum* example is a case in which intelligent activities like maze-solving are enacted through externally constructed niches that modify these brainless creatures’ activity (Crippen 2020, 2021; Crippen and Schulkin 2020, Ch., 5; Rolla & Figueiredo, 2021). Our goal, however, is not simply to reiterate the scaffolded mind claim, but to show how mindreading applies to the environment.

For now, we just want to stress the active and mind-scaffolding dynamics of surroundings, and accordingly offer another illustration with Salma, this time at an outdoor eatery, accompanied by her friend Moe, plus waiters, plastic tables, chairs and a cat sculking for food. These constituents delineate possibilities of action and organization, with each entity doing different things: Salma sips from her cup, and the table holds her yogurt smoothie and her resting elbows. Removing the table means Salma can traverse the space it occupied, but no longer rest her elbows or smoothie one meter above the ground. Limits also vary depending on the active centers that push and move various parts of the

environment. The cat is one active center, Salma another. Because the two come to the locale with different capacities, they face, as the expression goes, different situations. Hence, they perceive their situations differently—yet this is equivalent to perceiving and indeed being in objectively different situations. After all, the cat and Salma face objectively different constraints and enabling conditions, so that it is no mere subjective variation that differentiates the experience of sipping yogurt with a straw from that of lapping it while on all fours. Compared to the cat, Salma and Moe are embodied and consequently situated similarly and thus perceive things in comparable ways. Yet the two will nevertheless find themselves in very different situations if later travelling on a metro in which unwanted sexual stares and groping are common. This is a situation where Salma will confront objectively more risk than Moe and thus find herself in an emotionally threatening world, with certain gazes manifesting as aggressive to her in ways that they do not for Moe—again, because the objectively disparate situations of the two inflect expressions differently.

In sum, Salma may encounter the eatery as safe, the metro less so, perceiving and cognizing the latter as emotionally threatening, with affective tones especially punctuated and thereby mapped onto faces and bodies of others, all of this scaffolding experience. She may accordingly retreat from the subway station and get a cab. Gibson (1979, Ch. 8) advances a comparable idea when he suggests that a cliff-edge affords falling and looks threatening and is dangerous to most humans. But this is less so to skilled climbers with requisite equipment, who therefore do not value the cliff in the same way. Here, people see things differently, but primarily because the cliff really affords less danger to experienced individuals with the right tools. So the fact that two people read an environment differently does not refute the objective reality of what both perceive, even though the different perspectives are partly agent-generated. The same holds for faces.

Together, this suggests settings and faces are both multistable, such that a leer can make an environment more threatening, with surroundings doing the same for faces. However, unlike commentators who lean on examples of facial multistability never encountered in everyday contexts

and who explain the phenomenon within conventional representational frameworks (e.g., Boutet and Chaudhuri 2001; Block 2014), we have provided a 4E account and focused on ecologically valid cases. To clarify, our main stake is not about whether or not representations play a role at some point, and we do not argue this one way or the other. Instead, our primary claims are, first, that the multistability of faces reflects objective ecological changes that are explicable in Gibsonian and enactive terms, both of which insist on the palpable reality of what is observed; second, that reading expressions usually is not about decoding hidden emotions but about reading external situations; and, third, that faces and other things in the environment are parts of extended mental systems and also constituting structures of individual minds. A corollary that we pursue in our final section is this: if environments are themselves part of extended mental systems, then it may be that perceiving a setting's emotional tones is a genuine form of mindreading.

From Perceiving Emotions to Mindreading

Having considered how perceiving emotions is primarily a matter of registering situations, we now face a pressing problem: the challenge of specifying the putative relation between perceiving emotional expressions and mindreading, where the latter means having access to someone's mental states. At this juncture, we can endorse one of two theses. One option takes emotions to be fundamentally situational without going as far as to claim that perceiving them is a form of mindreading (see Crippen 2021). A reason for stopping at this weaker claim is the alternative would suggest the situated behavior of a robot such as a Roomba instantiates certain emotions or something functionally analogous, which may seem like a stretch. Further, if settings are emotionally laden, many will likewise find it counterintuitive to consider the scenes themselves as mindful.

On the other hand, a stronger thesis is that perceiving emotions is indeed a type of mindreading, which preserves the intuition that facial expressions give us access to mental states. Although this stronger view suggests itself from 4E standpoints and seems to hold the right results when we are

dealing with human social cognition, it invites potentially problematic counterexamples, especially in conjunction with the claim that emotions are situational. One of these is the seemingly puzzling idea that the complex behaviors of artificial agents as well as settings qualify as *bona fide* mindful. At first sight, this seems inimical to embodied accounts that take autonomy to be a fundamental criterion for mindedness (see Di Paolo et al. 2017). According to these views, a creature's open-ended activity of exchanging matter and energy with its surroundings in the process of preserving its organizational integrity is what allows for autonomy. For some embodied theorists like enactivists, biological agents uncontroversially meet the autonomy condition, thereby exhibiting natural purposes or intrinsic intentionality. Yet current artificial agents, on these same accounts, fail to do so—for their behavior is externally designed by others. In this section, we argue that perceiving emotions is indeed a matter of mindreading. Specifically, we advance a situated view in which psychic life is diffused into environments brought forth by their tenants rather than confined into discrete individuals. We thereby open the door to including non-biological entities within our scope, highlighting what this generally means to mindreading and social cognition and specifically for facial expressions.

When it comes to the epistemological question of how we can access other minds, traditional approaches such as theory-theory (Baron-Cohen 1995; Gopnik and Wellmann, 1992; Lewis, 1972) or simulationism (Gallese & Goldman 1998; Gordon 1986; Heal, 1986) agree that mindreading begins with the observation of others' behavior, where such behavior is taken to be the manifestation of hidden mentality. After this, however, the two positions diverge. Theory-theory relies on folk-psychology concepts in order to infer mental states from behavior. Simulationism holds that we access the mental states of others by implicitly running an internal virtual imitation (typically seen as correlated with our mirror neurons activating in tandem with observations of another's actions). One prominent alternative to both accounts is known as direct social perception (DSP), which has been championed by Gallagher (2008, 2017; and Varga 2014). DSP claims that accessing someone else's mind is not a matter of inferring the ulterior or hidden causes of their behavior from neutral

observations, nor a matter of representing through simulations, but consists in directly perceiving their mental states *in* their behavior (Krueger 2021). This fits an evolutionary picture where it is more important to see dangerous situations than impute private machinations upon encountering a rampaging bear when our retreat is cut off by a cliff edge, or when a predatory face probes us in a crime-ridden space. As Merleau-Ponty (1947/1964, pp. 52-53) concludes along these lines: “Anger, shame, hate, and love are not psychic facts hidden at the bottom of another’s consciousness: they are types of behavior or styles of conduct which are visible from the outside.”

A comparable rationale from DSP that builds on enactive and ecological outlooks is that, if the mind is embodied and if we directly register environmental possibilities for action, then to perceive someone’s face is to perceive at least in part their own mental states in a direct (non-inferential) manner. This position is strengthened by the fact that the ability to express emotions has been linked to the capacity to feel them and also forms a basis for social coordination and explicit communication (see Crippen and Schulkin, Ch. 3; Dumouchel 2022). In this regard, one’s mental and hence expressive state—say, of anger, curiosity, fear or longing—offers a distinct set of possible responses by an observer. Notice, however, that the idea of *observer* here is misleading and at best serves a useful heuristic, for we seldom find ourselves detached and passively observing others’ behavior (as in conventional models of social cognition). Instead, the way people act around us is shaped by our own doings, the way we react to them and, in some cases, by a shared history of our acting together. Consequently, social cognition involves interactors (De Jaegher 2009), who synchronize according to their own interpersonal dynamics (Di Paolo et al. 2018). Thus, when it comes to mindreading, the idea of interactors should replace neatly distinguished actors and observers.

The pragmatism at the heart of both enactive and ecological psychology (Heras-Escribano 2019; Crippen 2020) suggests that mindreading, like other forms of cognition, requires exploratory actions from the agent in their environment. Exploratory skill, in turn, entails selecting and enacting specific sensorimotor engagements afforded by the environment. In a similar fashion, an enactive-

ecological account of DSP is conceived as the selection and enaction of specific responses to an interactor's behavior, where such responses are enabled by engaging with them in specific settings. As Gallagher and Varga (2014, p. 189) state, "my perception of your action is already formed in terms of how I might respond to your action. I see your action, not as a fact that needs to be interpreted in terms of your mental states, but as a situated opportunity or affordance for my own action in response." We have seen this pattern before, as when Salma's comparatively more vulnerable embodied state brings out threatening facial expressions and a hostile environmental atmosphere because the objective risk of the setting leads her to respond more cautiously.

In social cognition, then, the *social* environment consists in "an agent in a pragmatic context that throws light on the intentions (or possible intentions) of that agent" (Gallagher 2014, p. 445), which might as well involve a cultural and institutional background. For example, the meanings—emotional and otherwise—of queuing depend on whether the line is for a bank or roller coaster. Once again, moreover, the general social mood is evident in the situation, with mundane dreariness infusing the first case and excited anticipation the second. Expressions, in their turn, inflect cultural contexts, so that jubilant vs. crestfallen faces at a football goal make for a very different situation, but either way this is all open to direct perception. Facial expressions can form parts of explicit narratives, for instance, by conveying "the reaction of an absent third party or to indicate agreement and affiliation with one present, or, on the contrary, to show disagreement and rejection" (Dumouchel 2022, n.p.). Here, faces enter inflective dialogues not only with the environing context but with one another, and function in what might be called communicative public forms of life, in Wittgenstein's (1953) sense.

In light of all we have said, it seems to be public situations we are reading when we respond to expressions, with facial behaviors a subspecies of this. When reading situations, moreover, we appear to be negotiating extended cognitive systems in ways paralleling *P. polycephalum*: both cases entail cognitively complex tasks, for instance, wayfinding and social communication; and both are only possible via the dynamics of group activity and scaffolding environments. Along these lines, Gallagher

(2008) mirrors his version of DPS after Gibson's (1966, 1979) ecological approach to the extent that the former, like the latter, argues agents can directly explore available information by acting. Gallagher elaborates the point, explaining that perception is here "smart" because its smartness exempts social cognition from the need to be enriched by further internal processes (e.g., inferences or simulations).

Recently, Overgaard (2017) has argued that the smartness required by Gallagher's DSP requires perceiving something *as* thus-and-so (Gallagher himself suggests so in his example of perceiving his car, 2008, p. 536). If true, this would indicate that internal computations are responsible for interpreting perception (perceiving *as*), thereby enriching social perception (see Block 2014). This is of course hostile to the anti-representationalism professed by most enactivists and other embodied thinkers. While we do not feel the need to dogmatically deny that representations have any role whatever in mental life, we do—as already stated—wish to defend a direct perception (and one could add a "direct cognition") account in Gibson's and Gallagher's vein.

A few points are worth noting in this regard. One is that our embodied capacity to respond in an interaction—which for Gibson and Gallagher means perceiving and cognizing—is a matter of know-how and readiness to act. On this view, complicated computations are rendered without representations, as when a fish tail solves fluid dynamics problems, or mountains funnel insects to breeding destinations, or again when Roomba's round body calculates spatial problems. Social cognition is similarly scaffolded by bodies in environments, which introduce normative constraints that are refined by successful practices with others (see Di Paolo et al. 2018; Myin and van den Herik 2020; Rolla and Huffermann 2021). Social cognition, therefore, depends on a practical grasp and hence ability to successfully respond to other people's solicitations and inhibitions, which are thereby made meaningful in enactive senses. If people clap and smile at a tear-streaked face, the meaning may be an individual is receiving an award, which the broader cultural and institutional contexts of an auditorium will make even more explicit. Moreover, virtually all adults will immediately grasp the situation and

unreflectingly align their behavior with the celebratory atmosphere and the various social responses it affords, with this in fact making the meaning manifest.

A second point carried in DSP theory is that the very idea of social know-how fits the everyday fact that we can refine other abilities with practice. Hence, growing intimacy facilitates reading someone else's mental states and intentions on her contextualized facial and bodily behavior. The same reasoning explains why, at first, we might struggle to interact if we find ourselves in unprecedented social circumstances, like having a conversation with someone brought up in a substantially different culture. The important thing here is that, to the extent that social know-how is crucial for DSP and that it involves practice and contextual sensitivity, it follows that social cognition is fundamentally situational.

Now, to return to the initial concern that, if emotions are primarily situational and if certain robots and settings can be said to display emotions, given that perceiving emotions is a form of mindreading, it follows that non-autonomous creatures and environments themselves are mindful. We admit this may seem counterintuitive, but we are willing to bite the bullet and accept that conclusion. In order to make it more appealing, a first step is to acknowledge that this may seem counterintuitive due to a lingering tendency to conceive ourselves as detached observers of the world around us. This is opposed to the view we have defended, which conceives agents as embedded and scaffolded interactors whose minds depend on the world with which they not only emmesh, but also enact. Road rage stands as another example where aggressive driving is an extension of angry looks and gestures, amplified by hostile responses from other drivers. All of this only happens in a road populated by multiple agents, together affording certain possibilities of action and restraint and comprising a cognitive system. This system pushes into and thereby animates or enlivens even concrete mediums and normally dead objects, which pose genuine threats like predators lying in wait, but only in this highly dynamic and active context. This explains not only why we may perceive non-living things *as if*

they have agency, but why indeed they do have agency in extended networks within which brains entangle with limbs and tools, along with impacts of surroundings that push back.

To some extent, a similar thing happens when we perceive a Roomba as if it was exploring the room and looking for dirt. Granted, we are not perceiving *its* mental state (given its mindlessness or lack of autonomy) but a distributed mental state that makes sense in an extended (i.e., non-immediate) environment which is also inhabited by engineers that have programmed the creature to behave so and consumers that set it to certain tasks. In other words, we do perceive a mental state typical of exploratory behavior directed at the purposeful action of dust gobbling, but it is distributed through the robot's chain of agents rather than tightly circumscribed by its physical boundaries. Thus, we would amend the assertions of Dumouchel (2022, n.p.), who characterizes such situations as “quite unlike what is the case in our relations with each other (and with some animals). In [human-robot interactions], there is no need to tend to the relation, pay heed to its evolution, be careful or nurture it.” We of course agree that robots are different than humans, but if some animals are admitted into the sphere of things that need nurturing, then why not a plant sadly wilting from lack of water? And if that, then it seems we can include as mindful a robot vacuum upsettingly choking on a cord, thereby struggling to fulfil purposes established via the network of extended mental intentions of which it is an actor.

Conclusion

An opening argument in this paper was that perceiving emotional expressions is situational. By this we not only meant to question the project of identifying a universal correspondence between specific facial behavior and emotional states, but also to suggest expression reading is primarily about assessing situations. From there, we developed a positive account according to which situations have emotional characters that get close to what Gibson designates as affordances, thereby pointing out that emotion-related (and mind-related) properties are most fundamentally specifiable in interactions. We

then moved on to the thesis that minds are situated, which highlights that not only bodily structure, but also other creatures and objects encountered in environments, limit and enable the way we can engage with things, all this scaffolding psychic life. Along the way, we used AI to elaborate on shortcomings in Ekman's project, for example, citing the unlikelihood of systems that can interpret either the Kuleshov effect or real-life analogues of situationally inflected expression that appear pervasively in the world. We also employed behavior-based robotics and AI to explicate situated cognition and as an example of why it is not, after all, misguided to see the non-animate environment as minded, making the same claim about other things such as city settings, winding trails and suchlike.

The above suggest that accurately perceiving someone else's behavior is not a matter of assuming a detached observational stance that provides inputs for internal representations about their mental states. Perceiving other people's emotions and their mental states is instead a matter of perceiving possibilities of interaction and responding to them, as enactivists have argued in their version of Direct Social Perception. But in contrast with other proponents of DSP, we take this view a step further in arguing that whole systems (such as environments and their populations) are parts of extended intentional systems and can therefore be perceived as having emotions of their own. We have done so by comparing emotional environmental colors to tones that arise in the context of situated facial behavior, arguing that these cases are not so different, thereby rejecting the idea that people over-impute mind to things.

References

- Aviezer H, Trope Y, Todorov A (2012) Holistic person processing: Faces with bodies tell the whole story. *Journal of Personality and Social Psychology*, 103: 20-37.
- Aviezer H, Yaacov T, Todorov A (2012) Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, 338 (6111), 1225-1229.
- Baron-Cohen S (1995) *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press.

- Barrett Feldman L, Adolphs R, Marsella S, Martinez A, Pollak S (2019) Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1), 1-68.
- Berridge K (2018) Evolving concepts of emotion and motivation. *Frontiers in Psychology* 9. <https://doi.org/10.3389/fpsyg.2018.01647>.
- Block N (2014) Seeing-as in the light of Vision Science. *Philosophy and Phenomenological Research*, 89, 560-572. <https://doi.org/10.1111/phpr.12135>
- Boden M (2006) *Mind as machine: A history of cognitive science*, Vol. 1 & 2. Oxford: Oxford University Press.
- Boutet I, Chaudhuri A (2001) Multistability of overlapped face stimuli is dependent upon orientation. *Perception*, 30, 743-753. <https://doi.org/10.1068/p3183>
- Brooks R (1990) Elephants don't play chess. *Robotics and Autonomous Systems*, 6: 3-15.
- Buck R (1994) Social and emotional functions in facial expression and communication: The readout hypothesis. *Biological Psychology* 38: 95-115.
- Collingwood RG (1939) *An Autobiography*. Oxford: Oxford University Press.
- Carroll JM, Russell, JA (1996) Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology* 70: 205-218.
- Carroll N (1993) Toward a theory of point-of-view editing: Communication, emotion, and the movies. *Poetics Today*, 14, 123-141.
- Chemero A.(2009) *Radical Embodied Cognitive Science*. The MIT Press.
- Clark A, Chalmers D (1998) The extended mind. *Analysis* 58: 7-19.
- Crawford, K (2021) *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.
- Crippen M (2016) Dewey, enactivism and Greek thought. In Madzia R and Jung M (eds), *Pragmatism and embodied cognitive science: From bodily interaction to symbolic articulation*, 229-246. Berlin: De Gruyter.
- Crippen, M. (2017). Embodied cognition and perception: Dewey, science and skepticism. *Contemporary Pragmatism* 14, 121-134.
- Crippen M (2019) Pragmatic evolutions of the Kantian a priori: From the mental to the bodily. In: Skowronski K, Pihlstrom S, (eds), *Pragmatist Kant*. Nordic Pragmatism Network, Helsinki, pp 150-171

- Crippen M (2021) “Aesthetics and Action: Situations, Emotional Perception and the Kuleshov Effect.” *Synthese* 198 (S9): 2345-2363. <https://doi.org/10.1007/s11229-019-02110-2>.
- Crippen M (2020) Enactive pragmatism and ecological psychology. *Frontiers in Psychology* 11. <https://doi.org/10.3389/fpsyg.2020.538644>.
- Crippen M, Schulkin J (2020). *Mind ecologies: Body, brain, and world*. New York: Columbia University Press.
- Crivelli C, Jarillo S, Fridlund AJ (2016) A multidisciplinary approach to research in small-scale societies: studying emotions and facial expression in the field. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2016.01073>
- De Jaegher H (2009) Social understanding through direct perception? Yes, by interacting. *Consciousness and Cognition*, 18, 535–542. <https://doi.org/10.1016/j.concog.2008.10.007>
- Dennett D (2006) *Breaking the spell: Religion as a natural phenomenon*. New York: Viking.
- Dewey J (1920) *Reconstruction in philosophy*. New York: Henry Holt and Company.
- Dewey J.(1929) *The quest for certainty*. New York: Minton, Balch and Company.
- Dewey J (1934). *Art as experience*. New York: Minton, Balch and Company.
- Di Paolo E, Buhrmann T, Barandiaran X (2017) *Sensorimotor life: An enactive proposal*. Oxford: Oxford University Press.
- Di Paolo E, Cuffari EC, De Jaegher H (2018). *Linguistic Bodies: The Continuity Between Life and Language*. MIT Press.
- Dumouchel P (2022) Making Faces. *Topoi*. <https://doi.org/10.1007/s11245-022-09799-w>
- Dreyfus H (1991). *Being-in-the-world: A commentary on Heidegger's Being and time*, division I. London: MIT Press.
- Dreyfus H (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Philosophical Psychology* 20: 247-268. <https://doi.org/10.1080/09515080701239510>
- Du S, Tao Y, Martinez AM (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences*, 111, E1454-E1462. <https://doi.org/10.1073/pnas.1322355111>
- Ekman P (1972). Universals and cultural differences in facial expressions of emotion. In: Cole JK (ed.) *Nebraska symposium on motivation* (vol. 19), 207-283). Lincoln: University of Nebraska Press.
- Ekman, P (1992) Are there basic emotions? *Psychological Review* 99: 550-553.

- Ekman P (2016) What scientists who study emotion agree about. *Perspectives on Psychological Science* 11: 31-34. <https://doi.org/10.1177/1745691615596992>.
- Ekman P, Friesen WV (2003) *Unmasking the face: A guide to recognizing emotions from facial clues*. Englewood Cliffs: Prentice-Hall.
- Ekman P, O’Sullivan M (1988). The role of context in interpreting facial expression: Comment on Russell and Fehr (1987). *Journal of Experimental Psychology: General*, 117, 86-88.
- Feigenbaum E (1968). Artificial intelligence: Themes in the second decade (memo 67). Stanford: Stanford University Intelligence Project.
- Feldman Barrett, L (2012). Emotions are real. *Emotion*, 12(3), 413-429. <https://doi.org/10.1037/a0027555>
- Feldman Barrett L, Gendron M, Huang Y (2009). Do discrete emotions exist? *Philosophical Psychology* 22: 427-437. <https://doi.org/10.1080/09515080903153634>
- Gallagher R, Ludovica J (2019). “We Tested Europe’s New Lie Detector for Travelers—and Immediately Triggered a False Positive.” *The Intercept*, July 26. <https://theintercept.com/2019/07/26/europe-border-control-ai-lie-detector/>.
- Gallagher S (2008). Understanding Others: Embodied Social Cognition. In P. Calvo & T. Gomila (Eds.), *Handbook of Cognitive Science: An Embodied Approach* (pp. 439–452). Elsevier.
- Gallagher S (2017). *Enactivist Interventions: Rethinking the Mind* (Vol. 1). Oxford University Press. <https://doi.org/10.1093/oso/9780198794325.001.0001>
- Gallagher S, Varga S (2014) Social Constraints on the Direct Perception of Emotions and Intentions. *Topoi*, 33: 195–199.
- Gallese V, Goldman A (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12), 493–501.
- Gibson JJ (1966) *The senses considered as perceptual systems*. Boston: Houghton-Mifflin.
- Gibson JJ (1979) *The ecological approach to visual perception*. Boston: Houghton-Mifflin.
- Gopnik A, Wellmann HM (1992). Why the Child’s theory of mind really is a theory. *Mind & Language*, 7(1–2), 145–171. <https://doi.org/10.1111/j.1468-0017.1992.tb00202.x>
- Gordon RM (1986) Folk psychology as simulation. *Mind & Language*, 1: 158–171. <https://doi.org/10.1111/j.1468-0017.1986.tb00324.x>
- Grey W (1950) An imitation of life. *Scientific American*: 182(5): 42-45.
- Gu S, Wang F, Patel NP, Bourgeois JA, Huang JH (2019). A model for basic emotions

using observations of behavior in drosophila. *Frontiers in Psychology*, 10.
<https://doi.org/10.3389/fpsyg.2019.00781>

Heal J (1986) Replication and functionalism. In J. Butterfield (Ed.), *Language, Mind and Logic*. Cambridge University Press.

Heaven D (2020). Why faces don't always tell the truth about feelings. *Nature* 578 (7796): 502-504.

Heras-Escribano M (2019). Pragmatism, enactivism, and ecological psychology: towards a unified approach to post-cognitivism. *Synthese*, 0123456789. <https://doi.org/10.1007/s11229-019-02111-1>

Hutto D, Myin E (2013) *Radicalizing enactivism: Basic minds without content*. MIT Press.

Jack R, Garrod O, Yu H, Caldara R, Schyns P (2012) Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences* 109:7241-7244. doi:10.1073/pnas.1200155109.

Jack R, Sun W, Delis I, Garrod O, Schyns P (2016). Four not six: Revealing culturally common facial expressions of emotion. *Journal of Experimental Psychology: General* 145: 708-730.

Koffka K (1935) *Principles of gestalt psychology*. New York: Harcourt, Brace and Company.

Köhler W (1947) *Gestalt psychology: An introduction to new concepts in modern psychology*. New York: Liveright.

Krueger J (2021) Enactivism, other minds, and mental disorders. *Synthese*, 198(S1), 365–389.
<https://doi.org/10.1007/s11229-019-02133-9>

Lewis D (1972) Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50(3), 249–258.

Long J (2011) *Darwin's devices: What evolving robots can teach us about the history of life and the future of technology*. New York: Basic Books.

Merleau-Ponty M (1945/1962) *Phenomenology of perception*. Trans: Smith C. New York: Routledge and Kegan Paul.

Merleau-Ponty M (1947/1964) Film and the new psychology. Trans: Dreyfus H, Dreyfus P
In *Sense and non-sense*, 43–59. Evanston: Northwestern University Press.

Myin E, van den Herik JC (2020) A twofold tale of one mind: revisiting REC's multi-storey story. *Synthese*. <https://doi.org/10.1007/s11229-020-02857-z>

Nakagaki T, Yamada H, Tóth Á (2000). Maze-solving by an amoeboid organism. *Nature*, 407(6803), 470–470. <https://doi.org/10.1038/35035159>

- Overgaard S (2017) Enactivism and the perception of others' emotions. *Midwest Studies In Philosophy*, 41: 105-129. <https://doi.org/10.1111/misp.12068>
- Pudovkin V (1926/1968) Film technique. In Montagu I (ed. & trans) *Film technique and film acting* London: Vision Press, pp. 19–220
- Rolla G, Figueiredo N (2021) Bringing forth a world, literally. *Phenomenology and the Cognitive Sciences*. <https://doi.org/10.1007/s11097-021-09760-z>
- Rolla G, Huffermann J (2021). Converging enactivisms: radical enactivism meets linguistic bodies. *Adaptive Behavior*. <https://doi.org/10.1177/10597123211020782>
- Rowlands, M. (2010). *The New Science of the Mind: From Extended Mind to Embodied Phenomenology*. The MIT Press.
- Russell J (1993) Forced-choice response format in the study of facial expression. *Motivation and Emotion*, 17, 41-51.
- Russell J (1994) Is there universal recognition of emotion from facial expression? A review of cross-cultural studies. *Psychological Bulletin*, 115, 102-141.
- Russell J (2006) Emotions are not modules. *Canadian Journal of Philosophy*, 36, 53-71.
- Simon H (1969) *The Sciences of the Artificial*. Cambridge, MA: MIT Press, 1969.
- Smith G (2018) *The AI delusion*. Oxford: Oxford University Press.
- Smith M, Cottrell G, Gosselin F, Schyns P (2005) Transmitting and decoding facial expressions. *Psychological Science*, 16, 184-189.
- Stadler M, Kruse P (1995) The function of meaning in cognitive order formation. In *Ambiguity in mind and nature: Multistable cognitive phenomena*. Berlin: Springer, 6-21.
- Sterelny K (2010) Minds: Extended or scaffolded? *Phenomenology and the Cognitive Sciences*, 9(4), 465–481. <https://doi.org/10.1007/s11097-010-9174-y>
- Thompson E (2004) Life and mind: From autopoiesis to neurophenomenology. A tribute to Francisco Varela. *Phenomenology and the Cognitive Sciences* 4, 381-398.
- Tomkins S, McCarter R (1964) What and where are the primary effects: Some evidence for a theory. *Perceptual and Motor Skills*, 18, 119-158.
- Varela F, Thompson E and Rosch E (1991). *The embodied mind*. Cambridge: MIT Press.
- Weinberger S (2010). Airport security: Intent to deceive. *Nature*, 465(7297), 412-415.
- Werner H (1927/1978). On physiognomic modes of perception and their experimental investigation. In Barten, Franklin MB (Eds), *Developmental processes: Heinz Werner's selected writings*

(Vol. 1, pp. 149–152). New York: International Universities Press.

Wilson, RA (2004) *Boundaries of the Mind*. Cambridge: Cambridge University Press.

Wittgenstein L (1953). *Philosophical investigations*. Trans. GEM. Anscombe. Oxford: Basil Blackwell, 1953.