Stephan Käufer and Anthony Chemero (2021). Phenomenology, 2nd edition, Polity Press.

12

Enactivism and the Embodied Mind

12.1 Embodied, embedded, extended, enactive

In the wake of Merleau-Ponty's phenomenological philosophy, Gibson's ecological psychology, and the critiques of artificial intelligence by Dreyfus, increasing numbers of philosophers of mind and cognitive scientists began to question the rationalist assumptions that were foundational in the cognitive sciences. What these three historical precursors to what are now called "enactivism" and "embodied mind" have in common is a recognition of the importance of the body and action in understanding our mental lives. They also share a skepticism about internal, mental representations. To paraphrase Alva Noë (2004), experience is not something that happens to our brains; instead, it is something we do, typically by moving our bodies around our physical and cultural environments.

Enactivism and the embodied mind took shape in the 1980s and early 1990s. Arguably the true founding of both embodied mind and enactivism comes with *The Embodied Mind* by Francisco Varela, Evan Thompson, and Eleanor Rosch (1991). This book fuses Merleau-Ponty's phenomenology and work in neuroscience and situated robotics. The key claim of the book is that cognition does not involve computation or internal representations, but is instead a basic activity of living things. This was a radical rejection of the predominant cognitive science of the time, which took cognition to be computational processing of representations in the brain. In the late 1990s, some enactive ideas were mainstreamed and combined with representational and computational cognitive science, especially by Andy Clark (1997; but see also Wilson 1995 and McClamrock 1995). In effect, Clark's version of the embodied mind is what is now called "embodied mind"; Varela et al.'s original version is called "enactivism."

In the twenty-first century, there has been a profusion of theories that fall under the general idea of embodied mind, so much so that it is currently fashionable to lump them together as "4E cognition" (Menary 2010), where the Es in question are embodied, embedded, extended, and enactive. The first three of these are variations on a theme; the fourth is something of an outlier. To say that mind is *embodied* is to say that cognitive systems include aspects of the body outside the nervous system; to say that it is *embedded* is to say that environmental, social, or cultural resources are necessary for at least some cognitive activities; to say that mind is *extended* is to say that those environmental resources are literally components of the cognitive system (Clark and Chalmers 1998; Kono 2010). Notice that each of these successive Es is a stronger claim. Notice too that each of these three Es is compatible with mainstream representational and computational theories of mind. Indeed, arguments for extended mind often take it to be a straightforward consequence of functionalism in the philosophy of mind (Drayson 2010; Wheeler 2010), in which things in the environment play crucial roles in the cognitive processes that constitute cognition. For example, instead of remembering phone numbers using neural systems, you might store them in a phone or write them down in a notebook. The notebook out in the world is part of your mind. In this sense the mind is extended. Nevertheless the notebook functions by storing and representing information and the extended mind uses this representation to cognize the world. So the extended mind is a representing, computing mind. Saying that mind is *enactive* is to say something quite different; it is to say that mind is a fundamental activity of living things. As noted above, enactivism is incompatible with representational and computational cognitive science. It is the only one of the four Es that is genuinely in line with phenomenological thinking.

12.2 The original enactivism

The roots of the enactive approach are in biology, in the work of Humberto Maturana and Francisco Varela (Maturana and Varela 1973, 1987; Varela, Maturana, and Uribe 1974). Maturana and Varela developed the theory of *autopoiesis*, an attempt to give a theory of life and

KAUFER & CHEMERO 9781509540655 PRINT.indd 224 05/01/2021 11:29

living systems that is amenable to mathematical and computational modeling. "Autopoiesis" translates from Greek as "self-creation." Maturana and Varela thought that being self-creating and self-maintaining is the key property of living things. In their analysis, autopoiesis has two components: autopoietic systems are *operationally closed* and *structurally coupled* to their environments. To be operationally closed is to be autonomous, in that a particular system's activities create and maintain those very activities. Suppose there is a set of two chemical reactions, which are such that a product of reaction A is a catalyst for reaction B and a product of B is a catalyst for reaction A. These two reactions could form an operationally closed system in that reaction A makes reaction B possible and vice versa. To be structurally coupled to one another, two entities must have a history of interactions that leads, over time, to a congruence between them. The key example of an autopoietic system that Maturana and Varela give is the cell. A cell is a set of chemical reactions that are bounded by a semi-permeable cell wall. The cell wall maintains the chemical reactions by keeping the concentrations of chemicals favorable; the chemical reactions in the cell create and maintain the cell wall. The wall and the reactions form an operationally closed set. The cell is structurally coupled to the extra-cellular environment in that the wall constantly and selectively admits raw materials for the chemical reactions from the extra-cellular environment, and constantly and selectively passes waste products to the extra-cellular environment. The cell impacts the chemical concentrations' extra-cellular environment, which also impacts the chemical concentrations in the cell. The cell, then, is an autopoietic system, structurally coupled to its extra-cellular environment. Maturana and Varela take autopoiesis to be the essential characteristic of living systems. More recently, Di Paolo (2008) has argued convincingly that autopoiesis is in itself not sufficient for life. What he calls *adaptivity* is also required. Adaptivity is the ability of a system to tell when it is approaching the boundaries of its viability, and to act so as to change its circumstances. This has been widely accepted as an important amendment to autopoietic theory. It shows, for example, why the Brooks robots discussed in the previous chapter are not alive.

This early work by Maturana and Varela is the basis for enactivism, which combines autopoiesis with insights from phenomenology. As we have seen, the work that is viewed as the founding document in enactivism is *The Embodied Mind* by Varela, Thompson, and Rosch (1991). That work chronicles the history of cognitive science, showing its repeated failures to capture human experience, and offers an alternative approach based on autopoietic theory, ideas from Merleau-Ponty, and the robots of Rodney Brooks described in Chapter 11. Although *The*

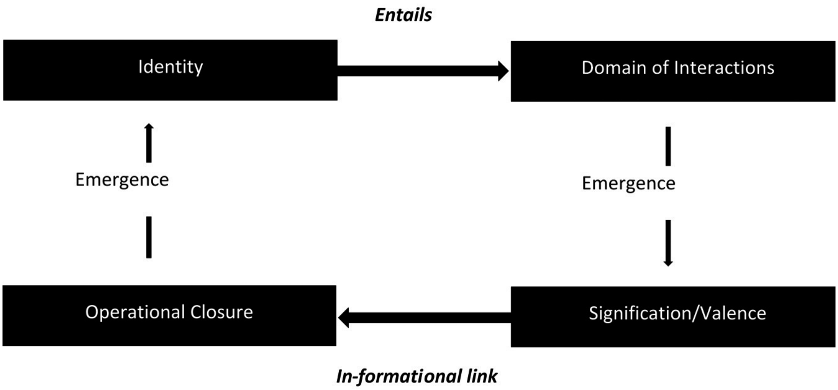

Figure 12.1 Co-emergence of self and world
Source: Based on Thompson 2004

Embodied Mind and *Mind in Life*, a later book by Thompson (2007), have been the most influential works in enactive cognitive science, the most clear and concise description of the first enactivist position is found in a 2004 article by Thompson, a tribute to Varela, who had died in 2001. Our exposition of enactivism in the next paragraph follows the exposition in that paper closely.

Thompson (2004) begins by pointing out that the key to understanding the relationship between experience and the material world, that is, to solving the mind–body problem, is what Merleau-Ponty called the lived body. The main question that enactive cognitive science attempts to answer concerns the relationship between a biological living body and a phenomenological lived body, the relationship between *Körper* and *Leib*. Figure 12.1 is based on a figure from Thompson's paper. The beginning point is that life is autopoiesis. Because autopoietic systems are operationally closed, there will be a separation of the living system from its environment, as in the case of the cell wall. This separation is what makes a living organism an entity separate from its environment. It also implies the emergence of a self, a primitive self in the case of a cell, but a self nonetheless. The emergence of the self implies the emergence of a world, not just in that the boundary around the self leaves everything else as the world, but also in that the activities of the living system pick out which aspects of the world it structurally couples with. This world and the self co-emerge, insofar as the activities of the autopoietic system determine what aspects of the physical environment it structurally couples to. Enactivists often call this co-emergence of self and world "sense-making" (Varela 1979; Thompson 2004; Thompson and Stapleton 2008), by which they mean that the experienced world for the organism is the sense it makes of its environment. Sense-making is both cognitive and emotional, so the world is significant to the organism; it has value and valence. Indeed, sense-making *is* cognition, at least in a minimal sense. In sense-making an organism maintains itself in a meaningful environment. This meaningful environment does not exist in advance of the existence of the organism, but co-emerges along with the activities of the organism. Organisms, as lived bodies, enact or "bring forth" worlds (Varela, Thompson, and Rosch 1991). This is now a biological, living body and also a phenomenological, lived body.

Varela, Thompson, and Rosch (1991) exemplify this bringing forth of worlds using the Rodney Brooks robot Allen, discussed in the previous chapter. As we noted there, Allen senses using twelve ultrasonic sensors positioned at each "hour" around its circular body. The only things, therefore, that can perturb Allen's behavior are physical objects large enough to reflect ultrasonic pulses, and the way Allen reacts to these objects determines their significance to Allen. This is what it is for Allen to enact or bring forth a very limited world of significance. This is taken as a model for all organisms, each of which brings forth a world that is determined by the connections between its sensing and acting, and these connections determine the significance of the entities in that world. The world that is experiences is already, automatically significant.

Enactivists are interested in extended mind. But because they focus on the bringing forth of worlds, extended mind takes on a different cast. Because organisms bring forth the world they experience, that world, in all its significance, is not independent of the organism, so it is strange to claim that the cognitive system extends beyond the organism. On the one hand, because the experienced world is not separable from the organism, mind could not but be extended; on the other hand, because the world is brought forth by the organism, it is strange to say that the world is external to the organism. Enactivists, therefore, focus on the possibility of extended life instead of extended mind. Di Paolo (2008) discusses the behavior of water boatmen, a species of insect that is able to breathe underwater by trapping air bubbles in abdominal hairs. Because of pressure differences that result from the boatman's own respiration, bubbles replenish themselves with oxygen, allowing for extended periods under water. These bubbles mediate environmental coupling, and alter the boatman's abilities to interact with its environment, and in so doing alter the significance of entities in its world. Di Paolo (2008) argues that these boatmen-plus-bubbles comprise an extended form of living. Thompson and Stapleton (2008) argue that cases like this are like the connection between Merleau-Ponty's blind man and his cane. Just as the blind man

KAUFER & CHEMERO 9781509540655 PRINT.indd 227 05/01/2021 11:29

does not experience his cane, the water boatman does not experience the bubble, but experiences an altered world of significance through the bubble. The bubble for the water boatman is equipment. As Merleau-Ponty claimed, the water boatman and cane navigator have *incorporated* their equipment, experiencing the world through that equipment. Since life is cognition according to these enactivists, extended life is a form of extended mind.

De Jaegher and Di Paolo (2007) extend enactivism to social interactions, via what they call *participatory sense-making*. In participatory sense-making, two individuals are coupled with the world and with one another, such that they collectively and temporarily open a new domain of significant interactions that is not available to either separately. To take an example from De Jaegher and Di Paolo, consider what happens in what we might call "the hallway dance," when each of two individual humans attempts to make space for the other while passing in a narrow hallway. Each of these individuals is an adaptive agent, engaging in sense-making, and bringing forth a significant world. Given their interest in avoiding collision with one another, which would have negative valence for both, each will move to one side of the hallway to let the other pass. Most of the time this works perfectly well, but sometimes both individuals will move to the same side of the hallway, yielding a potential collision. Because each experiences potential collisions negatively, both simultaneously then move to the opposite side of the narrow hallway, setting up another potential collision. To avoid the second possible collision, each individual once again switches sides of the hallway. This can happen several times, with each person mirroring the other's movement repeatedly. This hallway dance is an example of participatory sense-making. In it, each individual remains an autonomous agent, bringing forth a significant world, even while they are temporarily coupled with one another, but they also collectively bring forth a world significant to their coupled activity. Notice that, like the individuals that participate in it, the hallway dance displays its own autonomy (Fuchs and De Jaegher 2009). The dance maintains itself, at least for a while, because each attempt at collision avoidance by the participants leads to another possible collision, the avoidance of which leads to another possible collision, and so on.

In the case of the hallway dance, the participatory sense-making is at odds with what is adaptive for the individuals: neither of them wants to bounce from wall to wall in the narrow corridor, but they are temporarily trapped in a social interaction. This is not always the case in participatory sense-making, and much of our social interaction also serves the individuals in the interaction. Consider (non-hallway) dancing, conversation, or joint speech (such as protest or football chants; see Cummins 2018). In these cases, a new domain of significance, unavailable to the individual participants separately, is opened by the interaction, but, in this case, the participatory domain is positively valenced for the individuals. In these cases, the interaction maintains itself over time, for however long it does, because the individual participants (dancers, conversationalists) work to maintain the interaction.

In recent work, Cuffari, Di Paolo, and De Jaegher (2015; see also Di Paolo, Cuffari, and De Jaegher 2018) expand participatory sense-making into an account of language. Rather than trying to explain language as an abstract entity that humans use, they focus on what Maturana (1978) calls *languaging*, an activity central to our lives as human beings. Cuffari et al.'s full account of languaging is complex, and includes a dialectical and a developmental model. We will focus here on the latter. Beginning even before they are born, humans are enmeshed in a world that is first and foremost social, constituted by close interactions with care-givers. That is, the sense-making of infants is primarily participatory sense-making. The environment in which this occurs is replete with speech and gestures; the environment is "enlanguaged." During the early years of their lives, children develop skills at sense-making, alone and with others. Because this sense-making occurs in an enlanguaged environment, the skills and habits with which they make sense involve the use of language. That is, just as the water boatman has incorporated its bubble and the blind person has incorporated her cane, developing children incorporate pieces of language. As their sense-making habits and skills become more and more inflected with language, and their participatory sense-making with care-givers and a widening circle of others depends more and more on speaking and listening, children become what Cuffari et al. call "linguistic bodies."

From a phenomenological point of view, this is an attractive view of language and linguistic activities. Unlike cognitivist approaches to language, which focus on innate neural propensities for syntax and grammar, the enactive approach accounts for language as a skilled activity with which we make sense of the material, social, and broader cultural worlds. In this way the enactivist view of language fits with Heidegger's claim that "language, as the holistic totality of words in which discourse has its own 'worldly' being, shows up within the world just like available entities" (*SZ*, p. 161). Language is equipment and our linguistic abilities are practical competences for using it correctly. Moreover, again unlike the cognitive approach, the enactive approach is what Cuffari et al. call "nonrepresentational." In languaging, humans make sense of the world, but not typically by representing it. Cuffari et al.'s view differs from "antirepresentational" approaches because it acknowledges that one of the things we can do with language is make representations, as when we tell a story or compare a loved one to a summer's day.

12.3 Other enactivisms: The sensorimotor approach and radical enactivism

Although it is often also called "enactivism," the strong sensorimotor approach associated primarily with Alva Noë , J. Kevin O'Regan, and Susan Hurley (O'Regan and Noë 2001; Hurley 2002; Hurley and Noë 2003; Noë 2004, 2009: O'Regan 2011) is rather different from enactivism as described above. Although it is influenced by Merleau-Ponty, it is more closely related to Gibson's approach. To avoid confusion, we will call it the "sensorimotor approach." The key to the sensorimotor approach is that perceiving, seeing, experiencing, and the like are things that we *do*, not things that happen inside of us. Consider dynamic touch, described in Chapter 10. To be able to perceive by touch, you need to explore by actively hefting, running your fingers over edges, and the like. Indeed, Gibson (1962) showed that human participants could not identify objects that were pressed into their palms or run over the edges of their fingers by experimenters, but could identify them very precisely (e.g., "a snowman-shaped cookie cutter") when allowed to explore them with their hands. Perceiving by touch requires exploratory work, and sensory stimulation alone is not sufficient. According to the sensorimotor approach, all our senses are like touch in that they require active exploration of the world. Accordingly, this active exploration is part of experience, so the sensorimotor approach is a form of extended mind, with experiences depending on the brain, body, and environment.

The key differentiating feature of the sensorimotor approach is its explanatory use of *sensorimotor contingencies*. Sensorimotor contingencies are relationships between bodily movements and changes in sensory stimulation (O'Regan and Noë 2001). Sensorimotor contingencies are in some ways like Husserl's inner horizons, and in some ways like Gibson's affordances. To use an example from Noë (2004), when you look at a tomato, only a small portion of it is reflecting light that strikes you in the eyes. You nonetheless see it as a three-dimensional object and as having a back. The back of the tomato is present, even though light reflecting off it is not striking your eyes. You also see it as something you could touch or bite or slice. You see all of these things in virtue of an awareness of sensorimotor contingencies, how you would be stimulated if you acted in a particular way. You see the back because if you leaned forward, light from it would strike your eyes; you see that you could touch it because if you reached out with your hand, your finger would press on its flesh. Having experiences at all, according to the sensorimotor approach, requires awareness of these sensorimotor contingencies.

Among the consequences of the sensorimotor approach is a scientifically accessible approach to conscious experience (O'Regan 2011). The sensorimotor approach has a readily available explanation of why the different senses differ from one another. We can differentiate between, for example, vision and touch because they have different sets of sensorimotor contingencies. Moving my fingers over an object yields different changes in stimulation than moving my eyes over it will. This is buttressed by experimental research with sensory substitution. In the 1970s, Bach-y-Rita and colleagues began experimenting with connections between an eyeglass-mounted camera and arrays of vibrating motors worn on the back or stomach or, later, tongue (Bach-y-Rita and Kercel 2003). This is called tactile-vision sensory substitution or TVSS. With the eyeglass-mounted camera, explorations of the environment had vision-like sensorimotor contingencies: turning the head to the left, for example, led to changes in the light entering the camera, and so to changes in vibrations of the array of motors, that were similar to visual changes. Those wearing the TVSS system could quickly learn to identify objects and navigate around cluttered rooms. Moreover, they reported the feeling of *seeing* things around them, rather than feeling a pattern of vibrations on their tongues. This strongly suggests that the experiential difference between seeing and touch has nothing to do with the nature of the stimulation (light hitting photoreceptors in the eyes, vibrations on the skin), but is instead determined by the sensorimotor contingencies specific to each sense. This suggests that the sensory substitution device is part of the extended cognitive system (Auvray and Myin 2009), meaning that this version of enactivism also embraces the extended mind.

We consider yet another approach called "enactivism" primarily for the sake of completeness. What Daniel Hutto and Erik Myin (2013, 2017) call "radical enactivism" consists primarily in the rejection of representations in what they call "basic cognition." Basic cognition is cognition that is not informed by public language or other cultural symbol systems. Basic cognition, Hutto and Myin argue, involves no computation or representations. So, they argue, there are representations only in human linguistic activities, and not in perception, motor control, and the like. They group their arguments for this claim into what they call "don't need" and "can't have" arguments. For the former, they say that basic cognition can be explained just in terms of biological function and without referring to representations and computation, so theorists don't need representationalism. For the latter, they argue that the very idea of mental representations is incoherent, so theorists can't have them for their explanations.

We find many of these arguments convincing, and generally endorse Hutto and Myin's anti-representationalism about basic cognition. We are less convinced by their representationalism about language use and cultural cognition. The phenomenologists we have described throughout this book generally assume that, although language and cultural resources can be used to construct representations, language and culture are not representational at their base. As we have seen above, so too do the others described in this chapter who call themselves enactivists. Although they align themselves with other enactivists and Gibsonian ecological psychologists, Hutto and Myin's form of enactivism differs in that it is not inspired by phenomenology. This is, of course, not a criticism, but rather an explanation for this book's short discussion of radical enactivism.

12.4 Enactivism as a philosophy of nature

Enactivism is often called "enactive cognitive science." We have, so far, avoided that term, because enactivism is not, in and of itself, a scientific view. Indeed, in two recent major works in enactivism (Di Paolo, Buhrmann, and Barandiaran 2017; Gallagher 2017), it has been claimed that enactivism is best viewed not as a scientific research program but as a philosophy of nature. (This distinction comes from Godfrey-Smith [2001].) A scientific research program is a set of hypotheses, at least some of which are testable, and has methods and practices for testing some of those hypotheses, analyzing data, and evaluating results. In short, a scientific research program is what we generally call a "science"; examples include cognitive neuroscience, plate tectonics, and condensed matter physics. In contrast, a philosophy of nature is a philosophical stance about some features of the natural world and scientific attempts to make sense of them. A philosophy of nature is broader than an individual scientific research program, and develops stances toward the methods and results of multiple scientific research programs, interpreting and critiquing their results, and sometimes inspiring new scientific endeavors. Godfrey-Smith introduces the distinction to argue that developmental systems theory, a theoretical approach in biology that inspires and interprets research in scientific research programs such as evolutionary biology, developmental biology, heredity, and genetics, is a philosophy of nature and not a scientific research program. (See

Oyama, Griffiths, and Gray [2001] for more on developmental systems theory.)

Gallagher argues that enactivism should be seen as a philosophy of nature because "from the very start enactivism involved not only a rethinking of the nature of mind and brain, but also a rethinking of the concept of nature itself" (2017, p. 23). Indeed, as we have seen above, the first work in enactivism was a rethinking of the nature of living things, and of the relationship between life and mind. Moreover, enactivism has none of the trappings of a scientific research program. It is a philosophical position, in most cases inspired by the work of the phenomenological thinkers we have described in earlier chapters. As a philosophy of nature, enactivism has inspired work in scientific research programs. For example, as noted above, Maturana and Varela intended autopoiesis to be a theory of life amenable to mathematical and computational modeling. Given this, autopoiesis and enactivism are foundational in the field of *artificial life*. Long before desktop computers were common, Varela, Maturana, and Uribe (1974) had demonstrated an autopoietic system on a computer, building a virtual cell with a semi-permeable, self-repairing cell wall. Artificial life continues to be an important methodology for enactivist theorists (e.g., Froese and Di Paolo 2010; Egbert, Barandiaran, and Di Paolo 2011). So too does artificial intelligence: Froese and Ziemke (2009) describe an approach that they call "enactive artificial intelligence."

Enactivism is a philosophy of nature, and not a scientific research program. Arguably, the same is true about Gibson's ecological psychology described in Chapter 10. Even though he was employed as a psychology professor and did empirical science, in his *The Ecological Approach to Visual Perception*, Gibson told a detailed story about the nature of perception and action, and of the world in which animals perceive and act. That book contains no descriptions of testable hypotheses, experimental methods, or data analysis tools. Like enactivism, ecological psychology is more a philosophy of nature than a scientific research program. As good philosophies of nature, both ecological psychology and enactivism are inspired by and have inspired significant scientific research. In the next chapter, we will explore research programs in the cognitive sciences inspired by several of the phenomenologists we have discussed in this book.

Key terms

adaptivity – the ability of living systems to respond to situations in which they approach the limits of their viability.

autopoiesis – literally, self-creation. According to enactivists, autopoiesis is living and living is cognition.

embodied mind – a combination of the views of Gibson, Heidegger, and Merleau-Ponty with computational cognitive science. According to the theory of the embodied mind, cognition is a kind of computation in which some of the computational operations are done by using the body.

enactivism – a theory of cognition according to which cognition is the activity through which living things bring forth a significant world.

extended mind – the view that cognitive systems sometimes include portions of the non-biological environment.

operational closure –a key part of autopoiesis. A system is operationally closed when it maintains all its own operations.

participatory sense-making – a state in which two individuals are coupled with the world and with one another, such that they collectively and temporarily open a new domain of significant interactions that is not available to either individual separately.

sense-making – what an autopoietic system engages in when it finds significance in its world. This occurs even in simple systems that respond differentially to different situations.

sensorimotor contingencies – relationships between movements and changes in sensory stimulation that enable experience and allow for differentiation among the senses.

structural coupling – a state in which two systems' shared history leads to a congruence between them. Autopoietic systems are structurally coupled to their environments.

Further reading

Cuffari, E., E. Di Paolo, and H. De Jaegher (2015). "From participatory sense-making to language: There and back again." *Phenomenology and the Cognitive Sciences*, 14: 1089-1125.

Di Paolo, E., T. Buhrmann, and X. Barandiaran (2017). *Sensorimotor Life: An Enactive Proposal*. New York: Oxford University Press.

O'Regan, J. K. and A. Noë (2001). "A sensorimotor account of vision and visual consciousness." *Behavioral and Brain Sciences*, 24(5): 883–917.

Thompson E. (2004). "Life and mind: From autopoiesis to neurophenomenology. A tribute to Francisco Varela." *Phenomenology and Cognitive Science*, 3: 381–98.

Varela, F., E. Thompson, and E. Rosch (1991). *The Embodied Mind*. Cambridge, MA: MIT Press.