6

Perception without Inferences

In *Chapter 5* I looked at the notion of representation in action; in this chapter I'll look at the question of inference in perception. Are there inferences involved in perception, and if so, what does that mean, and where should we locate them? Here I'll join several others in arguing against an inference model of perception (e.g., Bruineberg, Kiverstein, and Rietveld 2016; Orlandi 2014) and I'll defend an enactivist alternative. I'll explore how the enactivist approach can respond to issues related to cognitive penetration and the effects of culture on perception. This approach, however, will raise some challenges in regard to how we might pursue a science of perception. Accordingly, I'll end by returning to considerations mentioned in *Chapter 1*, concerning enactivism understood as a philosophy of nature.

6.1 Inference Models of Perception

There is a long tradition in which perception is understood to involve inference. One can think immediately of Helmholtz, but also the more recently developed models of predictive coding. In between, but seemingly consistent with both of these approaches, we find the computationalist view, expressed, for example, by Fodor and Pylyshyn in their 'Establishment' critique of Gibson's notion of direct perception.

> The current Establishment theory (sometimes referred to as the 'information processing' view) is that perception depends, in several respects . . . upon inferences. Since inference is a process in which premises are displayed and consequences derived, and since that takes time, it is part and parcel of the information processing view that there is an intrinsic connection between perception and memory. And since, finally, the Establishment holds that the psychological

> mechanism of inference is the transformation of mental representations, it follows that perception is in relevant respects a computational process.
>
> (Fodor and Pylyshyn 1981, 139–40)

One question is whether we should take the notion of inference literally or metaphorically (see Hatfield 2002 for review). For example, Helmholtz contends that the processes of perception 'are *like inferences* insofar as we from the observed effect on our senses arrive at an idea of the cause of this effect' (Helmholtz 1867, 430). Likewise, Palmer, expressing what Fodor and Pylyshyn describe as the Establishment view, states:

> Using the term 'inference' to describe such a process may seem to be somewhat metaphorical and thus to undercut the force of the claim that perception works by unconscious inference. But, as we said at the outset, unconscious inference must be at least somewhat metaphorical, since normal inference is quite clearly slow, laborious, and conscious, whereas perception is fast, easy, and unconscious. The important point for present purposes is that perception relies on processes that can be usefully viewed as inferences that require heuristic assumptions.
>
> (Palmer 1999, 83).

Finally, in the predictive coding camp, Jakob Hohwy sometimes makes the same gesture.

> The problem of perception is the problem of using the effects—that is, the sensory data that is all the brain has access to—to figure out the causes. It is then a problem of causal inference for the brain, *analogous* in many respects to our everyday reasoning about cause and effect, and to scientific methods of causal inference . . . There is a sense in which, in spite of being Bayesian, [prediction error minimization] is more mechanical than it is inferential . . . The 'neuronal hardware' of the mechanism itself is not literally inferential: neuronal populations are just trying to generate activity that anticipates their input.
>
> (Hohwy 2013, 13, 55)

The pervasive claim that brain processes are best understood as inferences, if sometimes regarded as metaphorical, is nonetheless often treated as a substantive claim. If the claim is that perceptual processes operate *as if* they were inferential (computational, representational), it would be difficult to disagree: many things can be viewed as if they operated inferentially—thermostats, smoke detectors, etc. But this would mean that it's just one way of characterizing the subpersonal brain processes of perception, and if we can find a different and simpler, i.e., more parsimonious, way of characterizing such processes, or one that has equal or greater explanatory power, we should consider it as a viable alternative.

What motivates the idea that brain processes are inferential?

1. A general unobservability principle (UP).

The brain has no direct access to the world. Helmholtz (1867, 430) expressed this well: 'We always in fact only have direct access to the events at the nerves, that is, we sense the effects, never the external objects'. Or as Jacob Hohwy puts it, 'the sensory data . . . is all the brain has access to' (2013, 13). Predictive coders, including Hohwy, should in fact deny the last proposition since they would also argue that the brain has access to a set of priors (memory), based on prior experience. But this is only further explication of why they think perception is inferential, since inference seems to be the best mechanism to explain how priors, which may be conceptual in format, get integrated with sensory information. A more specialized version of UP motivates the idea that social cognition involves theoretical inference—because what we want to know (the other's mental state) is hidden from us. Just as we have no direct access to the other person's mind, our own brain, more generally, has no direct access to the world.

The assumptions are clear. First, assume that all cognition (including perception) happens in the brain. Second, assume that the brain is locked up in the darkness of the skull and has no access to the outside world. It's a mystery how it could gain knowledge of current worldly affairs without drawing inferences from the clues provided by sensory data—the only clues it seemingly has available. Andy Clark summarizes:

> For, the task of the brain, when viewed from a certain distance, can seem impossible: it must discover information about the likely causes of impinging signals without any form of direct access to their source . . . [A]ll that it 'knows', in any direct sense, are the ways its own states (e.g., spike trains) flow and alter. In that (restricted) sense, all the system has direct access to is its own states. The world itself is thus off-limits (though [it] can, importantly, issue motor commands and await developments) . . . How, simply on the basis of patterns of changes in its own internal states, is it to alter and adapt its responses so as to tune itself to act as a useful node (one that merits its relatively huge metabolic expense) for the origination of adaptive responses? . . . The task is . . . to infer the nature of the signal source (the world) from just the varying input signal itself.
>
> (Clark 2013a, 183)

Because the brain is isolated from the world—locked up in the skull—we are led to ascribe a complex structure or procedure involving computations, inferences, and representations, a structured process that helps the system (the brain) work out a solution.

This is precisely the predictive coding view. The brain is pictured as having no direct access to the outside world; accordingly, it needs to represent that world by some internal model that it constructs by decoding sensory input (Hohwy 2013). This process involves synaptic inhibition based on empirical priors. Based on priors (i.e., memories, assumptions, or prior experiences) and given a certain sensory input, the brain is pictured as making top-down probabilistic inferences about the causes of that input. Predictions are then matched against ongoing sensory input. Mismatches generate prediction errors that are sent back up the line; the brain corrects for those errors, and the system adjusts dynamically back and forth until there is a relatively good fit.

2. A second motivation for the inference explanation is that it offers sufficient explanatory power to resolve the poverty of stimulus problem.

This is based on the assumption that the stimulus does not provide all the information needed for perception—the so-called poverty of stimulus problem. For example, the pattern of light that hits the retina is ambiguous—compatible with the production of a variety of visual representations of the world. Note that this again depends on (1), i.e., the brain has no access to the world.

> This fact predicts that we should see the world in a radically unstable way: but this is patently false. The inferential view explains stability by supposing that the visual system makes use of some stored assumptions that help reduce the representations produced in response to the retinal stimulus. (Orlandi 2013, 743)

On the predictive coding model, priors, based on Bayesian statistical processes, feed and constrain a process of inference (prediction) formation. For example, visual light discontinuities might be caused by various environmental features—edges, cracks, shadows, etc. But since edges are statistically the more frequent cause of light discontinuities, that is the prior that informs the inference. Unless that inference generates prediction errors, the brain settles in a stable way on this being a perception of an edge.

6.2 An Example from Social Cognition

Whether inference is a necessary ingredient of perception has been the subject of recent debates about direct (i.e., non-inferential) perception in the context of social cognition. There has been, in fact, a surprising turn in this debate. Some proponents of theory of mind (ToM), a position that has traditionally held to the Unobservability Principle (UP), have given up this principle and have embraced the idea of direct social perception for some mental states (e.g. Carruthers 2015; Lavelle 2012). UP has been seemingly one of the central assumptions leading to the idea that one requires inference for ToM processes, precisely because, as traditionally held, we do not have perceptual access to the other's mind. Alan Leslie (of many who could be cited) gives clear expression of this position: 'Because the mental states of others (and indeed ourselves) are completely hidden from the senses, they can only ever be inferred' (Leslie 2004, 164). In giving up UP it might seem that we should give up the idea that we need inference to understand others, and that our knowledge of other minds via perception could be non-inferential. But that's not how it turns out.

Carruthers (2015, 3), for instance, argues 'the mental states of other people are often represented *in the content of perception* [and] this conclusion is consistent with many forms of theory-theory [inferential] accounts of our mindreading abilities'. It remains consistent with ToM because, according to Carruthers, perception itself is shaped by a tacit set of ToM inferences.

> Indeed, one might think that *any* adequate account of our perception of mental states (assuming that the latter is real) would need to appeal to a set of tacit inferences underlying such perceptions, which might then qualify as a form of theory-theory. (Carruthers 2015, 2)

Inferences are required because mental states cannot be perceived independently of concepts and acquired knowledge of the world. 'We have no idea how to explain the causal processes involved except by appealing to something amounting to a tacit *theory*, I suggest . . . even enactivism cannot obviate the need for tacit theory' (Carruthers 2015, 3).

In conceiving of perception in this context, there are at least three possibilities.

1. We need perceptual processes plus extra-perceptual inferential processes; the inferences are conceptual/theoretical and the integration happens so fast it *seems* to be purely perceptual (this is the position defended by Lavelle and Carruthers).
2. Perception is itself an inferential process (Helmholtz and the predictive coding view); it is not that we add a cognitive process to perceptual processes; perception is already cognitive.
3. Perception is enactive (action-oriented, affordance driven, and not inferential), but is nonetheless epistemic and 'smart' because it is attuned to context and can take direction.

Lavelle (2012), for example, makes it clear that to grasp another person's mental states (e.g., intentions, emotions), perception must be supplemented by extra-perceptual inferential processes. This does not occur on the phenomenological level of our conscious experiences (i.e., on the integrative timescale), but subpersonally (on the elemental timescale). We need not add an extra step to our conscious experience; but that's because the extra (inferential) step is already in the mix—already added at the subpersonal level.

Lavelle rejects, for example, Gallese's proposal that low-level mirror neuron processes are sufficient for understanding actions. Gallese had argued that '[W]e don't need to suppose an over-arching top-down influence in order to have a neural mechanism that maps the goal. We already have it in the premotor [or parietal] system. We don't need to imply a further mechanism that maps the goal' (Gallese 2007, 15). In contrast, Lavelle sticks to the idea that beliefs are represented in non-perceptual areas and are introduced via inferential processes. 'The moral is that while theoretical entities need not be unobservable, one requires a theory in order to observe them' (2012, 228).

Carruthers (2015), in agreement with Lavelle, still wants to distinguish between perception and inference, but also suggests that perception is never '*encapsulated* from the remainder of one's beliefs and goals. This way of thinking of the perception/cognition boundary presumes that there is a stage in visual processing—sometimes called "early vision"—that is beyond any direct influence from one's other mental states . . . it is doubtful whether there is any such stage' (2015, 3). While this almost suggests a perceptual process that is fully integrated with the conceptual, in a way that would not require inference—a perception that is cognitively penetrated *simpliciter*—Carruthers still wants to retain the inference process. He regards it as a fast, online process.

> Acquiring concepts that classify a set of arbitrary similar-seeming shapes into two distinct categories, for example, transforms the perceived similarity spaces among the shapes. Those that seemed similar before now seem distinctively different as a result of category acquisition. Until recently, however, it was unclear to what extent these effects reflect a late decision-like stage in processing, or whether sensory experience is altered by concepts in an online manner. But there is now ample evidence of the latter. (Carruthers 2015, 5)

Another example that he cites suggests the same thing. The Greek language has two words for 'blue' (light blue [*ghalazio*] vs dark blue [*ble*]), but only one for 'green'. Thierry et al. (2009), using EEG, measured a specific neuronal signal, *visual mismatch negativity* (vMMN) over the visual cortex, in Greek and English speakers. This signal occurs *c.*200 milliseconds following the presentation of an oddball stimulus (for example, a square in a series of circles, or a dark blue circle in a series of light blue circles). On this elemental timescale, this signal is pre-attentive and an unconscious stage of visual processing. For blue contrasts, vMMN showed significant difference in Greek speakers, but no significant difference for English speakers. That is, the conceptual difference between different shades of blue shows up directly in visual processing. Thus, Carruthers concludes: 'Because Greek speakers have distinct *concepts* for light blue and dark blue, they see the two colors as more unlike one another and they do so from quite early stages in visual processing, prior to the impact of attention or judgment' (2015, 5).

It's not clear that these examples settle the question completely since we can still ask whether the influence on early visual processing could not be the result of plastic changes to early sensory areas. Carruthers, however, prefers a speedy online binding process between strict sensory/perceptual processing and culturally related concepts—something that still leaves room for unconscious inferences infecting the perceptual process. Carruthers doesn't mention that Thierry et al. (2009) also show differences between Greek and English speakers in P1 (the first, earliest positive peak elicited by visual stimuli over parietooccipital regions of the scalp at 100–130 milliseconds (100 milliseconds prior to vMMN). Given this, the inferential integration would have to be extremely speedy.

For Carruthers, however, the point is that *via* speed of processing, perceptual and conceptual processes are integrated by the time perception reaches consciousness.

> Conceptual information is processed within the window of a few hundred milliseconds that elapses between the presentation of a stimulus and its subsequent global broadcast. This could well be a function of expertise. While you or I might be capable of slowly figuring out, from the configuration of pieces on a chess board, that White has a winning position, a chess grandmaster may immediately *see* it as such... [In the case of social cognition, the] only limit will be whether *mindreading inferences* can be drawn fast enough for binding to take place. Since many forms of mental-state awareness are seemingly simultaneous with awareness of the behavior and/or circumstances that cause them, we can presume that ordinary mindreaders *can* draw the requisite inferences fast enough. (2015, 6–7)

In that case, even if I say that I see that you are upset, or that you intend to take a sip from your glass, this seeing (which on a conscious level seems direct) is really the result of subpersonally inferring your emotional state or intention on the basis of some perceptual cues and some basic rules of folk psychology. So, for Carruthers, this is still perception *plus* fast (indeed, very fast at 100 milliseconds if, following Thierry et al., we consider the earliest positive peak) theoretical inference, at the subpersonal level where a tacit ToM operates. What Lavelle and Carruthers suggest, then, is that perception and tacit theory are two separate things that need to be combined in quick inferential processes on the elemental timescale. This is not perception that is inferential, as in predictive coding (PC), but perception *plus* extra-perceptual inference.

Accordingly, this type of proposal contrasts with the PC or Helmholzian model, which considers perception itself to be inferential. Moreover, in PC models, as we noted, there is a more basic unobservability principle at work. This is the Helmholtzian idea, mentioned above, that the brain only has direct access to the events 'at the nerves', and never to the perceptual objects. The implication is that perception is not really direct. Rather, 'perceptual phenomenology [is] at one remove from the world... Interspersed between you and the world is a process of using models to explain away sensory input [i.e., to resolve prediction errors]' (Hohwy 2013, 48). This is 'a kind of indirectness'. Perception 'is indirect in the sense that what you experience *now* is given in your top-down prediction of your ongoing sensory input, rather than the bottom-up signal from the states of affairs themselves' (48).

6.3 The Enactivist Alternative

Enactivist approaches start with different assumptions and suggest a different vocabulary. The human brain not only evolved along with the human body, and works the way it does because of that; it's also not isolated, but rather is dynamically coupled to a body that is dynamically coupled to an environment. The organism (the brain–body system) is operating within the situation itself rather than on a model of the situation inferred by the brain. This coupling of brain–body–environment is structured by the physical aspects of neuronal processes, bodily movements, affects, anatomy and function, and environmental regularities.

Co-variance is physical variance across all parameters—brain, body, environment. For example, an object such as a piece of food in the near environment activates not only the visual cortex, but also other sensory areas, and the premotor and gustatory cortexes the way it does because I have hands and the object is reachable, and there is motivating interest (hunger) and an anticipation of reward. Change in any of these factors means that perception changes. If the human body evolved without hands, if I were aplasic (born without hands), or even if I lost my hands to amputation, to different degrees the whole system would perform differently and my perception would be different.

On the enactivist view, neural plasticity mitigates to some degree the need to think that subpersonal processes are inferential. The neural networks of perception are set up by previous experience—'set up to be set off', to borrow a phrase from Jesse Prinz (2004, 55). Whatever plastic changes have been effected in the visual cortex, or in the perceptual network constituted by early sensory and association areas, such changes constrain and shape the current response to visual stimuli. Neural networks are attuned to situated environmental events. Consider, in addition, that networked patterns of neuronal activation in premotor, parietal, and limbic areas modulate the dynamics of visual processing (Kranczioch et al. 2005; see Engel 2010, 233–5 for discussion).

For example, the limitations of my reach are determined not by my brain's representations of my arm, but by the physical length of my arm, which has grown as I have grown to adulthood, and to which my visual system has already physically accommodated itself. What this means is not that there is a fact or assumption about my arm length stored somewhere acting as a prior, and activated in an inference when the brain has to decide whether it's possible to reach the food. Rather, it means that there are physical changes that have occurred in visual cortical areas and in their connections with premotor and motor areas, so that when a hand-to-food-to-mouth action is called for, the system is activated in a dynamical causal fashion—it has already been set up to be set off in the right way. The physical length of my arm, which changes over the developmental timescale, together with my prior reaching practices, tune brain processes so that neuronal activations, rather than inferring anything, are attuning to my embodied physical possibilities and the physical affordances in a particular environment where something is either within reach or not, graspable or not, of interest or not, etc. I perceive things in terms of these sensory–motor contingencies and in terms of what those things pragmatically afford in relation to a body like mine, in the situation, also defined in part, for example, by my gustatory and more general interest conditions.

Many accounts of perception restrict the analysis to questions of recognition. As we've seen, the question in PC is often about how the visual system recognizes what is out there in the world, given that its access is limited to sensory input. This leads to the idea that the function of perception is to solve a puzzle, and what better way to solve a puzzle than to use inferential logic. But perception's function is never purely recognitional; vision, for example, involves more than recognition and motor control. The senses are not charged with just identifying or recognizing objects or just guiding bodily movement in the world. Response involves more than that; there are always ulterior motives. Because the organism desires food or rest or sex or aesthetic enjoyment or understanding, etc. the eye is never innocent. Consider that neuronal activity in the earliest of perceptual processing areas, such as V1, reflects more than simple feature detection. V1 neurons anticipate reward if they have been tuned by prior experience (Shuler and Bear 2006). This is not sensory data first, followed by inferential processes that conclude to reward possibility (an additional neural or cognitive function added on to sensory activation)—it's an intrinsically reward-oriented response or attunement to stimuli due to prior experiences and plastic changes—there's no room for or need for inferences in this respect. Perception is already attuned to reward possibilities.

Furthermore, with perception, autonomic and peripheral nervous systems are activated in dynamic patterns in synchrony with central processes—but in a way that makes it unclear what is regulating what. As we noted in section 1.5, along with the earliest visual processing, the medial orbital frontal cortex is activated initiating a train of muscular and hormonal changes throughout the body, modulating processes in organs, muscles, and joints associated with prior experience, and integrated with current sensory information (Barrett and Bar 2009). Just such modulations help to guide affective and action response. Integrated with affective and sensory–motor processes tied to the current situation, visual stimulation generates not just brain activation, but also specific bodily affective changes. Consider further that perception of another's face activates not just the face recognition area and ventral stream, but, importantly, the dorsal visual pathway that informs our motor system—suggesting that we perceive action affordances in the face of the other (Debruille, Brodeur, and Porras 2012). That is, we don't simply perceive the snapshot of a face in an instant with the task of recognizing it, we perceptually *respond* over time to affordances offered by the others' emotions as well as their actions.

Face perception presents not just objective patterns that we might recognize conceptually as emotions. It involves complex interactive behavioral and response patterns arising out of an active engagement with the other's face—not simple recognition of objective features, but interactive perception that constitutes an experience of significance or valence that shapes response. Social perception is affective in ways different from object perception. The experience of the gaze of another person directed back at you *affects* you, and your perception of the other's emotion *affects* you, even if this affect is not consciously recognized. Even when presented with masked, subliminal images of angry or happy faces or bodies, one's autonomic and peripheral systems register the emotion and respond (Tamietto 2013), and this response is part of what the perception is, as Barrett and Bar (2009) suggest. The perception of emotion is itself affective.

Is this best described as an inference process, perhaps a hierarchically organized set of inferential steps, a complex syllogistic argument that loops through the body to reach the conclusion that I'm attracted to or repulsed by what I see? It's one thing to think that the best way to talk about conceptual factors or folk psychological rules having an influence on perception is in terms of inferential processes, whether that is meant in a metaphorical or literal way. Likewise, if the only thing perceptually at stake for the organism were the task of recognition—of some object, a face, or another person—then it might seem intuitive to think that inference might do the job. But we (or our brains) are not simply trying to solve a puzzle that involves guessing or recognizing what is outside the skull. It seems less intuitive to think that broadly affective factors—including various autonomic and homeostatic factors—fit neatly into an inferential structure. Even a strong proponent of predictive coding, like Jakob Hohwy, can have doubts. '[T]he Bayesian, inferential approach to perception . . . seems rather intellectualistic . . . There is also something slightly odd about saying that the brain "infers", or "believes" things. In what sense does the brain know Bayes, if we don't? For that matter, a Bayesian approach to perception does not seem to directly concern the full richness of perceptual phenomenology . . . ' (Hohwy 2013, 18–19).

We've already seen (in *Chapter 2*) that it was impossible to fit all sorts of unruly bodily processes into a set of B-formatted representations, and we'll see (in *Chapter 8*) that all kinds of affective processes, and even variations in circulation and heartbeat, can influence perception (Garfinkel et al. 2014). Add to this the fact that respiration is not simply artifactual for perception but a causal factor contributing to the variability of neuronal responses to sensory stimuli and behavioral performance. We know, for example, that the obstruction or willful interruption of breathing increases cortical neuronal activity in sensory, motor, limbic, and association areas (e.g., Peiffer et al. 2008). It's also the case that MEG-measured changes in beta, delta, and theta oscillations and ongoing gamma power modulations in somatosensory cortex are driven by corresponding respiratory phase (a causal link via sensory input from the olfactory bulb), with the result that reaction time to a visual stimulus changes significantly, taking longer during inspiration than during the resting phase (Liu, Papanicolaou, and Heck 2014; see Heck et al. 2016).[1]

Such things as affects and the effects of respiration and heart rate are not *represented* as part of my perception; they are non-representational

[1] There are many other relevant effects that are tied to respiration; it affects movement and perceptual tracking (Rassler 2000); visual and auditory signal detection (Flexman, Demaree, and Simpson 1974; Li, Park, and Borg, 2012); emotion perception (Zelano et al., 2016); and pain perception (Iwabe, Ozaki, and Hashizume 2014; Zautra et al. 2010). Thanks to Somogy Varga for these references.

factors that have an effect on perceptual response. I (the experiencing agent) see what's in the world when my eyes are open, light is present, chemical changes happen on my retina, there's neural activation in visual cortex that connects to neural activation in premotor cortex and other areas that loop through affective, peripheral, autonomic, and other fully embodied systems. On the enactivist view, the perceptual system is not just in the brain; it includes the organism (brain–body) embedded in or engaged with an environment that is characterized by certain regularities and affordances and action possibilities. Take away some of the oxygen in the air and the entire system is affected—respiration, heart rate, digestion, postural balance, motor control. These are some symptoms of altitude sickness—which can also include double vision and irrational behavior.

The poverty of stimulus problem—the second motivation for the inference model—is addressed by the possibility of bodily movement—reflected in what enactivists call sensory–motor contingencies, and predictive coders call 'active inference'—moving around the environment provides more information and reduces the ambiguity. The point in such action is that the environment specifies itself (the environment is what it is)—*it* is not impoverished; the poverty only arises if we think that the brain has no access to the rich structure of the environment. It disappears if we acknowledge that the organism has access—is attuned or coupled—to the environment over time and is not only capable of movement, but is almost always moving. One can manufacture the poverty of stimulus problem by staying as still as possible—but this requires either some work or falling asleep.

Nico Orlandi has a nice argument in this context. The inferential view suggests that a default assumption informing the visual system is that edges are more statistically common than shadows or perhaps more evolutionarily important to detect. Given sensory stimulation by light discontinuity, the system goes to this default and forms the inference that it is perceiving an edge. In other words, to explain why the visual system infers what it does, the inferentialists must appeal to external environmental facts. Orlandi asks, 'why not do this from the get-go? If edges are more common or more typical or more important than other entities, then *that* is why we see them' (2014, 41). The visual system does not require an inference since, given evolutionary pressure or experience-driven plasticity, it 'can simply be wired by the environmental fact in question to produce states that track edges when exposed to discontinuities' (41). The system is physically attuned to such things, 'set up to be set off' by such visual discontinuities.

Just as there is more in the environment 'than meets the eye', there is also more in the environment that *does* meet the eye; that is, more than simple sensory cues such as visual discontinuities or colors. Visual cues taken together form a visual context, and integrated inter-modally with other sense modalities, including interoceptive proprioceptive, vestibular factors, etc. they capture a richer embodied and environmental context. This fact mitigates the need to think that what we call cognitive penetration works by fast knowledge-based inferences. If my visual system is wired to track edges, it should also be expected that, given the evolutionary and developmental significance of social relations, my perceptual system is (or comes to be) wired to track patterns of facial expressions, postures, movements, vocal intonations, and so forth. The tracking, which just is the co-variant activation of my perceptual system, is entirely perceptual, and the pattern that is perceived just is a sufficient part of the other person's emotion or intention to allow us to say that, without intervention of folk psychological inferences even at the subpersonal level, we perceive the emotion or intention (see Gallagher 2008b; Newen, Welpinghus, and Juckel 2015).

Along this line Orlandi (2014, 192ff.) shows, for the visual system, how context-sensitivity rather than knowledge-based inference can explain why we see a banana as more yellow than it actually is, or see African-American faces as darker than Caucasian faces even when they are exactly the same skin color (Levin and Banaji 2006). In the latter case, for example, it is not because we know factually that one face is African-American and infer that skin color must be comparatively darker, and so see it as such, but because we never see skin color all by itself; we also see shapes of noses and mouths—elements of a face pattern that we associate with darker skin color because of statistical regularity. If it were the result of knowledge-based inference, then we would expect the perceived difference to disappear once we knew that the skin color was identical—but it doesn't. As Orlandi goes on to show, one can explain such typical examples of cognitive penetration by reference to context sensitivity or semantic effects on attention—which is to say that these are not necessarily examples of cognitive penetration as this phenomenon is conceived by the inferentialists (also see Firestone and Scholl 2015).

Neural processes, coupled with non-neural processes, may co-vary with environmental regularities—certain light dispersals, certain tasty chemical patterns, certain textures, certain sound waves that the organism comes into contact with—but there's no reason to think of this in representational terms, or to think that such contact requires the mediation of inference making. Inferences are unmotivated if the organism (brain–body) is thought of as having access to, being attuning to, or being dynamically coupled to the environment. Attunement means that the organism is sensitive to certain environmental features—in part for evolutionary reasons (see *Chapter 9* for a more complex story), but also for reasons tied to ontogenetic development as well as to metaplastic effects of social and cultural factors. Such attuned sensitivity of organism to environment (shaped by reward patterns and affect patterns tied to learned responses and developed skills, which are themselves shaped by bodily details and environmental affordances) is sufficient for the perception of a significant world without requiring inferential processes seemingly tasked with constructing hypotheses about a world it cannot access.

6.4 Cultural Penetration

What is it that penetrates perception? Top-down, cognitive assumptions or beliefs? ToMish, folk psychological concepts or platitudes? Or some broader features of human (social) life? Moods, affects, traits, practices, and skills also can modulate perception (Siegel 2011). Some of these involve cultural factors. Perceiving others is not constrained simply by abstract differences in emotion patterns, but by situated affective attitudes towards out-group (*versus* in-group) members (Gutsell and Inzlicht 2010). Whether a person is able to respond to the emotions and intentions of another is crucially dependent on the person's attitudes (often implicit and nonconscious) about the racial or ethnic group to which the other belongs.

Evidence for this can be found in cross-cultural experiments on the perception of pain in others. A study by Xu et al. (2009) dramatically demonstrates the neural effects of implicit racial bias and shows that empathic neural responses to the other person's pain are modulated by the racial in-group/out-group relationship. fMRI brain imaging showed significant decreased activation in the anterior cingulate cortex (ACC), an area thought to correlate with empathic response, when subjects (Caucasians or Chinese) viewed racial out-group members (Chinese or Caucasian respectively) undergoing painful stimulations (needle penetration) to the face, compared to ACC activation when they viewed the same stimulations applied to racial in-group members. Differences in attitudes and biases are shaped by social and cultural experiences that likely cause plastic changes in ACC and SMA, and that clearly have some connectivity to our sensory systems. We are simply less responsive to out-group members and we display significantly less motor cortex activity when observing out-group members (Molnar-Szakacs et al. 2007). Most strikingly, in-group members fail to understand out-group member actions, and this is particularly prominent for disliked and dehumanized out-groups. The more dehumanized the out-group is, the less intuitive the grasp of out-group member intentions and actions (Gutsell and Inzlicht 2010).

On the inferentialist view, it is best to think of social and cultural factors in terms of theory-laden perception, as if the way our experience is (in)formed by social and cultural factors translates into the possession of a theory (a knowledge in the form of folk psychological beliefs or platitudes) that needs to be added to perception to formulate an extra inferential step in understanding others. The frequent example in discussions of cognitive penetrability involves beliefs. When you know that bananas are yellow, this knowledge affects what color you see bananas to be, so that an achromatic banana will appear to be yellow (Gegenfurtner, Olkkonen, and Walter 2006). This leads too quickly to the idea that perceptions are 'theory laden', a concept borrowed loosely from philosophy of science. Inferentialists will cite research to show that beliefs, and especially negative beliefs, about out-group members can interfere with one's ability to recognize emotions (Gutsell and Inzlicht 2010). In these cases making sense of the emotions of others is not constrained by cultural differences in emotion patterns, but by specific beliefs about the out-group member. On some conceptions it is not just a matter of 'having a belief' but of having a set of beliefs or a set of platitudes about the out-group that constitute part of folk psychology. On this view, the kind of subpersonal inferential processes suggested by Lavelle (2012) and Carruthers (2015) seem a possible explanation.

In contrast, rather than adding extra-perceptual inferential processing to perception, there is good evidence that perceptual processes at the subpersonal level are already shaped, via mechanisms of plasticity, by bodily and environmental (including social and cultural) factors and prior experience. For example, consider the now well-known difference between the way Westerners and Asians perceive and attend to visual objects and contexts (Goh and Park 2009). Westerners pay more attention to individual objects, while East Asians have an attentional bias toward backgrounds. These differences are not about the effects of particular beliefs or pieces of knowledge; they're regarded as differences in cognitive style that correlate not only with culture but also with age, and involve differences in the ventral visual areas of the brain (Goh et al. 2007). One also finds, for example, not only brain processes that are different relative to the use of different cultural tools and practices, but also cultural variations in brain mechanisms specifically underlying person perception and emotion regulation (Kitayama and Park 2010). For example, relative to European Americans, Asians show different neural processing in response to images of faces that represent a social-evaluative threat (Park and Kitayama 2014). In very specific ways, social and cultural factors have a physical, plastic effect on brain processes that shape basic perceptual experience and emotional responses.

This can help to explain why individuals are more accurate at recognizing the intentions and the emotions of members of their own culture versus those of other cultures (Elfenbein and Ambady 2002a, 2002b; Matsumoto 2002). There are subtle differences in emotional 'dialects' (or embodied interactional dynamics) across cultures, which reduce cross-cultural emotion recognition (Elfenbein et al. 2007). Research also shows that the in-group advantage in emotion recognition is largely independent of genetic or ethnic factors. It seems that individuals make best sense of emotions expressed by a member's own cultural group, regardless of race and ethnicity (Elfenbein and Ambady 2003). Again, however, if we regard emotion perception not simply as objective identification or intellectual recognition of the other's emotion, but as itself an embodied and affective process for the perceiver, as discussed in the previous section, it's not clear that theoretical inference will be sufficient to explain these phenomena. Moreover, theorists, like Carruthers, who defend the notion of innate, modular ToM mechanisms, or those who defend preprogrammed mirror systems operating in an automatic and context-independent fashion, have an especially difficult time explaining cultural differences in perception. For example, Scholl and Leslie (1999, 136–7) leave no room at all for these types of cultural effects, which may involve not only plastic changes in the brain but also metaplastic changes across the brain–body–environment system.

> One hallmark of the development of a modular cognitive capacity is that the end-state of the capacity is often strikingly uniform across individuals. Although the particulars of environmental interaction may affect the precise timetable with which the modular capacity manifests itself, what is eventually manifested is largely identical for all individuals. As the modular account thus predicts, the acquisition of ToM is largely uniform across both individuals and cultures. The essential character of ToM a person develops does not seem to depend on the character of their environment at all. It is at least plausible, prima facie, that we all have the same basic ToM! . . . The point is that the development of beliefs about beliefs seems remarkably uniform and stable.

Others maintain that the pattern of ToM development is identical across a species (e.g., Segal 1996), which is in marked contrast to the uneven and culturally dependent development of many other capacities. Again, however, cross-cultural studies of social cognition (see Domínguez et al. 2009 for a summary) and results from studies of racial bias and dehumanization (see Gallagher and Varga 2014) are inconsistent with these expectations, and show that mechanisms of social cognition are constitutively dependent upon historical-cultural situatedness and group membership. This suggests that the fundamental perceptual level of understanding others as persons is essentially dependent on cultural context—an aspect that any theory of social cognition must account for.

To deny that cultural factors have such effects on social perception, or perception in general, would only make sense if one were to accept the thesis of the 'cognitive impenetrability of perception' (Pylyshyn 1999). Both cognitive impenetrability and cognitive penetrability, however, conceive of the problem in the same way, because they conceive of the cognitive in the same way. In both cases, the cognitive is considered something that is stored on the upper floors of the brain, and then either inferentially injected, or not, into early perceptual areas of the brain. It's as if developmental and learning processes had an effect only on prefrontal or higher association areas and somehow passed through perceptual and motor areas without lasting effect. The effect only comes later when, in the case of cognitive penetration, there is 'just in time' delivery to effect the fast integration of conceptual information with sensory input.

In contrast, however, cultural aspects seem more pervasive. Not only beliefs, but also moods, traits, practices, and skills can modulate perception. For example, to the newly trained reader of Russian, a sheet of Cyrillic script looks different than it looked to her before she had the skill to read it (Siegel 2011). As Siegel points out, penetrated perceptions are confirmatory of the mood, trait, skill, etc. in a way that reinforces such things and can be epistemically pernicious. In fact there is a continuity of perniciousness from the cultural to the neurological. Cultural biases can reinforce neuronal firing patterns and result in plastic changes, reinforcing embodied practices and postures, behavioral habits, and intersubjective interactions. On the enactivist view, however, none of this counts against the idea that my perception of another's intentions and emotions is direct, requiring no extra-perceptual inference that would take us beyond what we perceive. All such changes, pernicious or not, are not additions to perception, an added-on set of inferences; rather, they transform the perceptual process itself.

6.5 Rethinking Nature: From Free-Energy to Autopoiesis

I return now to a theme I raised in section 1.6. Enactivist approaches present a challenge for science. As I indicated there, by focusing on not just the brain, not just the environment, not just behavior, but on the rich dynamics of brain–body–environment, where environment includes social and cultural factors, enactivism offers a holistic conception of cognition that is difficult to operationalize. Various practices and institutional arrangements that seem essential to good science—experimental controls, divisions of labor, disciplinary divisions—prevent taking all factors into consideration at once.

What would it mean to take up the distinction, suggested by Peter Godfrey-Smith (2001), between a 'scientific research programme' and a 'philosophy of nature', placing enactivism on the side of a philosophy of nature. As Godfrey-Smith suggests, a philosophy of nature would not necessarily have to share the same vocabulary as science. It 'can use its own categories and concepts, concepts developed for the task of describing the world as accurately as possible when a range of scientific descriptions are to be taken into account, and when a philosophical concern with the underlying structure of theories is appropriate' (Godfrey-Smith 2001, 284).

For a philosophy of nature to take scientific data seriously does not require that it take any particular scientific interpretation as necessary truth. Neuroscientific data, for example, tell us that certain neurons activate under certain circumstances. The open question is whether we have to interpret such activations in the standard scientific vocabulary of representations, inferences, simulations, computations, etc., all such interpretations predicated on a major internalist assumption (all cognition happens in the brain) that flies in the face of other data that point to the ongoing dynamical integration of such activations with bodily (affective, peripheral, and autonomic) and environmental (physical, social, and cultural) processes.

Being a pragmatist about the vocabulary of representation (as suggested in section 5.6), or about the vocabulary of inference, is at best only a temporary stance toward a set of placeholders that need ultimately to be cashed out not just in a different conception of brain function, but in a different philosophy of nature. An alternative way of thinking about nature should push hard on cognitive scientific practice in a way that makes doing science more difficult, but also more productive.

In this regard, enactivism involves not only a rethinking of the nature of mind and brain, but also a rethinking of the concept of nature itself (see Di Paolo 2005; Thompson 2007, 78ff.). Rethinking nature, as well as the nature of cognition, perception, and action, in terms of a continuity and integration of dynamical self-organizing adaptive systems where the distinction between physical and mental is deconstructed, where nature is not conceived purely in terms of objectivity, devoid of subjectivity, may further motivate a rethinking of science. As Daniel Hutto indicates,

> [E]nactivism is committed to the idea that mentality is something that emerges from the autopoietic, self-organizing and self-creating, activities of organisms. The activities in question are themselves thought of as essentially embedded and embodied interactions between organisms and their environments, interactions that occur and are themselves shaped in new ways over time. (2011, 22)

This transformation of the explanatory unit from brain to brain–body–environment is central to the challenge that faces the sciences of the mind. On one reading of predictive coding, the model of Bayesian inference entails a strong epistemic boundary that divides the brain from the rest of the body and the world (Hohwy 2013). On a different reading, the free-energy principle, which seems foundational for the PC approach as developed by Friston, points to a broader theoretical framework that links up with the concept of autopoiesis, which, in turn, plays a similar role for enactivism (Bruineberg, Kiverstein, and Rietveld 2016; Gallagher and Allen 2016; see Clark 2013a). Thinking in these broader terms affords a way to move closer to a philosophy of nature, or at least to a theoretical biology that might allow new insight into the more specialized problems of cognitive science.

The free-energy principle applies to any biological system that resists a tendency to disorder (Friston, Kilner, and Harrison 2006). It states that for an adaptive self-organizing (i.e., autopoietic) system to maintain itself it needs to minimize entropy or free-energy (or in PC terms, prediction errors). Variational free-energy, a mathematical concept, is, roughly, an information theoretic measure (the upper bound) of disorder or surprise. In theoretical biological terms, if we think of the living organism as a self-organizing system, it survives by anticipating sensory input or by taking action, which in turn changes its sensory input. Accordingly it needs to be attuned to its ecological niche in such a way that it minimizes surprise and 'the coupled dynamics of the organism-environment system remain within a relatively small subset of states that maintain the organism's viability in its econiche' (Bruineberg, Kiverstein, and Rietveld 2016, 2; see Friston 2011). Living systems and cognitive systems share this same organizational principle.

One important question is: how precisely does any particular system accomplish this minimization of surprise; how does it keep itself within the viability zone? Predictive coding is one possible answer, and, as we've seen, this particular answer is framed in terms of the Helmholtzian notion of unconscious inference, prediction error minimization, and active inference. This answer cuts the brain off from the world, however, at least theoretically. The brain has access only to its own processes and it has to predict its way to viability. The Markov blanket diagram represents this situation (Figure 6.1). The Markov blanket is a concept derived from formal treatments of Bayesian networks and causal dependency.

A Markov blanket defines a network or ensemble of nodes or subsystems that are interconnected by local deterministic forces. Markov blankets thus behave in a fashion similar to a cell wall, separating internal and external states to create stable dynamics that do not themselves directly

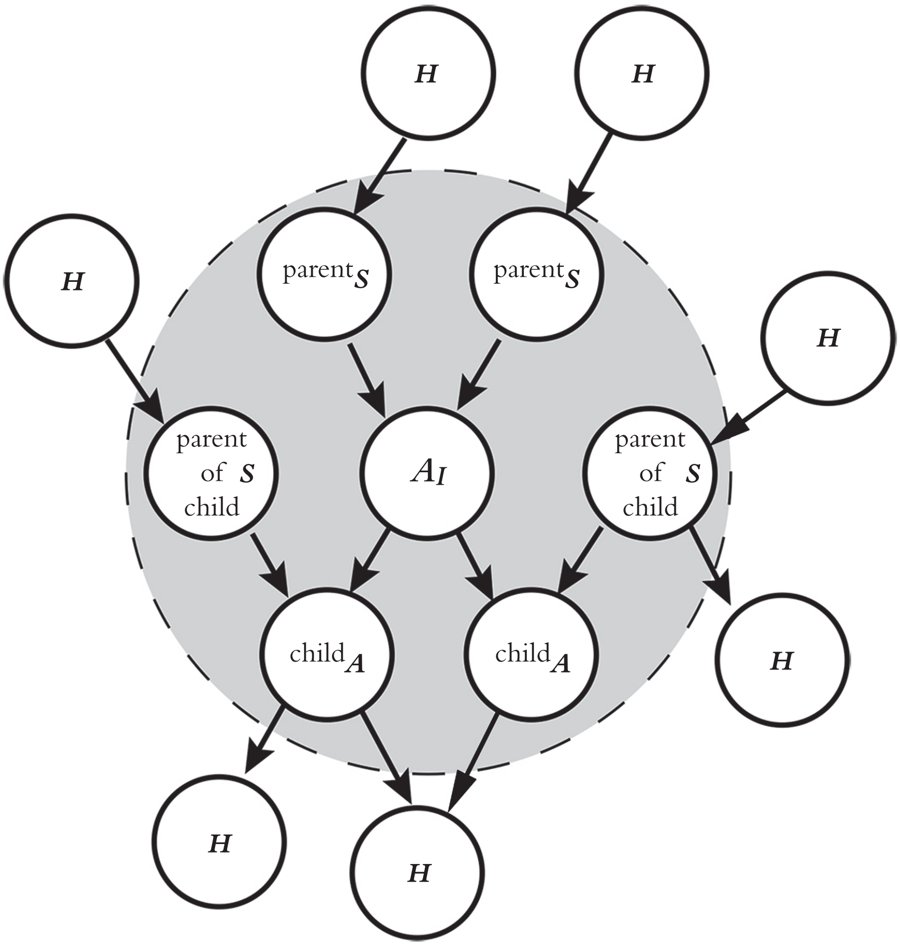

Figure 6.1. Markov blanket. The circle shaded in gray represents the Markov blanket of Node *A*, consisting of A, its children, parents, and parents of children, with parent/child being understood in terms of cause/effect. In small script, the sub-partition of internal and external states according to the Free Energy Principle (Friston 2013a); *H*, hidden external states, *I* internal states, *A* internal active states, and *S* internal sensory states. (Figure by M. Allen, from Gallagher and Allen 2016.)

impinge upon the local coupling responsible for their emergence. On the predictive coding interpretation, states external to the blanket can only be known indirectly (*via* inference). The blanket thus constitutes a partition between the world and the organism. Internal states themselves can be subdivided into those that are either the children of external states (sensations) or children of internal states (actions). Internal states have the capacity to probabilistically represent or infer hidden (external) states, which are themselves influenced by changes in the environment caused by active states. The inherent circularity of this scheme means that actions (which cause changes in the external world, but not sensory states) place an upper bound on the entropy of biological states, serving to maintain a homeostatic equilibrium constrained by internal states (Friston 2013a).

The predictive coding interpretation imposes a vocabulary of inference and representation on these dynamical, co-varying biological processes. As Bruineberg, Kiverstein, and Rietveld (2016) suggest, however, Friston's emphasis on circular causality and active inference leaves open another possible interpretation—the enactivist account, which emphasizes dynamical coupling of the organism with its environment, and works out the free-energy principle in terms of autopoiesis. In an autopoietic system, the boundary, represented in the Markov blanket, does not cut the system off from its environment but defines a coupling of organism–environment. 'The importance of such a boundary for living organisms has been central in the autopoietic approach from the very start . . . If this is the only kind of boundary that stems from the free-energy principle, then there seems to be nothing in the idea of probabilistic inference per se that challenges enactive cognitive science' (Bruineberg, Kiverstein, and Rietveld 2016, 22; see Bruineberg and Rietveld, 2014; Clark, 2015).

The organism, to minimize entropy, maintains internal homeostasis by hovering around an equilibrium point that keeps a balance among the conditions necessary for its viability and survival. It needs to minimize any surprises by anticipating possible threats and taking action to avoid or reduce them. Bruineberg, Kiverstein, and Rietveld (2016), citing Dewey's notion of organism–environment, mention an example that shows the importance of biological factors: 'For whales, being in deep sea is an event with low *surprisal*, and being on shore has high *surprisal*, while this is reversed for humans. Hence, the particular embodiment or biological organization of an animal and the environmental conditions of the animal necessary for its viability constrain each other' (p. 6). This general conception can be specified in terms of defining the basic (survival-enhancing) affordances that are relative to each animal.

In section 3.3 I pointed to the importance of Dewey's notion of situation, where situation is not equivalent to the objective environment, but includes the agent or experiencing subject in such a way that there is no way for the agent to gain an objective perspective on the situation. This is reflected in the idea that by perception alone the organism doesn't know or control its own viability conditions. It discovers them and can control them only by taking action.[2]

> So within the free-energy framework, it is *action* that does the work of actually minimizing surprisal. Actions change an organism's relation to the environment, thereby changing the sensory states of the organism, a process that Friston calls active inference (2012). Free energy, as we understand it, is a measure of the dis-attunement of the internal dynamics and the environmental dynamics [= surprise = increasing entropy]: it is low when the sensory states are anticipated, by the animal, and high when they are not. The free energy principle says that minimizing free-energy is a necessary and sufficient condition for living systems to maintain their organization in their econiche.
>
> (Bruineberg, Kiverstein, and Rietveld 2016, 9)

The enactivist interpretation places the emphasis on action; the PC interpretation places the emphasis on perception (Friston endorses both, noting that by perception we can also minimize surprise by adjusting our priors). Note, however, that action is not something happening in the brain, and is not just providing new sensory input for the brain; it's what the whole organism does in its interactions with the environment or, under a different description, what a person does in the world (see section 7.4), and this changes the world as much as it changes the brain. On this view, the priors that inform action are not assumptions or beliefs that inform inferences (as in PC); they're embodied skills and patterns of action-readiness that mesh with an affordance space (Brincker 2014; Gallagher 2015; see section 9.4). Perception is not isolated from such action; 'perception is an inevitable consequence of active exchange with the environment' (Friston 2009, 293). Perception and action involve dynamical adjustments to physical, social, and cultural affordances defined in terms of organism–environment, frequently involving normative practices that sometimes include science itself.

If nature cannot be understood apart from the finite cognitive capacities and action affordances that humans have to investigate it (and this is not only the enactivist view but a hermeneutical principle), this makes the scientific enterprise—which is itself a form of active engagement and exploration—more complicated. An enactivist philosophy of nature supports a kind of holism in which a plurality of factors are understood

[2] Friston (2010; 2013b) expresses this by saying that an agent does not *have* a model of its world, but rather *is* a model of its world.

to contribute to the full conception of mind. This is still a practical complication for experimental science, although it is certainly not necessary that in every case we must include absolutely everything when dealing with a particular concrete question. At the same time, practically speaking, it's a matter of deciding what factors are crucial, on the supposition that it may be easier to include than to ignore them. Including embodied interactions in explanations of social cognition, for example, might actually involve less complexity if keeping them out of the picture requires the elaboration of more convoluted explanations in terms of representations, inferences, or other concepts that may hold no other status than that of dominant metaphors. Likewise, including action, embodied processes, and environmental factors in explanations of perception and cognition more generally should lead us to a more comprehensive picture that is only as complex as it needs to be to account for the actual complexity of cognition.