Chapter 8

Virtual actions and abstract attitudes

> … each bodily stimulation for the normal subject awakens, not an actual movement, but a sort of "virtual movement"; the part of the body addressed escapes from anonymity, appears through a strange tension, and as a certain power for action within the frame of the anatomical apparatus. The normal subject's body is not merely ready to be mobilized by real situations that draw it toward themselves, it can also turn away from the world, apply its activity to the stimuli that are inscribed upon its sensory surfaces, lend itself to experiments and, more generally, be situated in the virtual. […] for the normal person each motor or tactile event gives rise in consciousness to an abundance of intentions that run from the body as a center of virtual action either toward the body itself or toward the object.
>
> —*Maurice Merleau-Ponty (1945/2012, p. 111)*

8.1 What next?

As we have stated at the beginning of this book, our goal is to put on the table a series of theoretically rich proposals for understanding embodied action and perception. We want to show that it is possible to move beyond the predominantly critical stance toward representationalism and articulate, in complementary fashion, a series of positive ideas and see them at work. We have not dwelled too long on debating representationalism apart from our brief discussion in Chapter 2. Many others have done this job before, and the debates will surely continue. Our sights are rather set on the promise that a world-involving, enactive story can go far without representations in any of the broad senses used by philosophers and cognitive scientists. We think this goal can be achieved even if, in its vagueness, the word representation can be made to cover a lot of ground indeed.

We also think that the foregoing chapters deliver on this promise. They draw on enactive ideas, dynamical formalizations, modeling work, empirical evidence, and phenomenological insights to support, expand, and complement existing embodied accounts of action and perception and give them a world-involving interpretation, one in which agent and environment co-define themselves through a history of engagements that cannot be captured in terms of input/output informational exchanges. The sensorimotor approach to perceptual experience—one of the most developed and most debated of these embodied accounts—has served us as a guiding thread. We have demonstrated that the notions of sensorimotor contingencies on which this account is based can be made precise in dynamical systems terms and that the concept of mastery admits powerful

non-representational interpretations. We have also responded to the worries raised by Thompson (2005) concerning the lack of a suitable account of subjectivity in sensorimotor contingencies theory by providing a compatible theory of agency and showing that it leads to a solid alternative account of the sense of agency.

Based on this, should we now expect an avalanche of embodied functionalists jumping ship to join the enactivist camp? Chances are that, despite the ground covered, in the mind of many we have not yet addressed the more pressing dilemmas facing non-representational accounts of action and perception. In part, this may be a typical "moving-target" situation. We should remember that many people still find it difficult to conceive of a non-representational account of mastery, let alone one of norms and goals as one should expect from a sound theory of agency, and that we have provided these here. But it is true that our account still needs to provide further answers to be more convincing; in particular, we should be able to decide whether it is feasible to explain more sophisticated forms of action with this approach. This feasibility, often denied to non-representational accounts, is the topic of this final chapter.

As we have seen in Chapter 2, representatives of embodied functionalism recognize the importance of the situated body and the benefits of dynamical explanations but they do not think these are sufficient for changing the general conception of cognition as a form of computation. They have felt secure all along because they rely on what they see as a stronghold beyond the reach of embodied/dynamical explanations: cases of so-called decoupled or "offline" cognition. At least some intelligent activity will not be explainable in terms of dynamical couplings with the environment because they demand the processing of information that is unavailable at the moment in the coupling.

What is wrong with this picture? We summarize some points already made in Chapter 2:

1. This is a sort of I-cannot-imagine-things-being-otherwise argument, an argument by default. These hardly ever provide a strong basis for deciding on scientific issues, especially in areas where our knowledge and conceptual categories are still rather limited and in flux. And by definition, it is impossible to question the adequacy of a theoretical framework by relying on these arguments.
2. Many criticisms of computationalism cast doubt on the information processing metaphor (e.g., the assumptions of near-decomposability of the human brain, the assumption of universal applicability of informational descriptions, the problems and limitations of a functionalist account of meaning, and so on). One cannot simply assume that certain experiences and mental phenomena (e.g., imagining unicorns) suffice to overthrow these criticisms, especially since in themselves these critical perspectives never deny those experiences (and in some cases they rely precisely on paying closer and more systematic attention to lived experience). In short, embodied functionalism is not problem-free and cannot be said to have successfully addressed the issues its defenders feel so confident it can explain.
3. The evidence of representational capacities in the human mind (imagination, symbolic thinking, mental imagery, etc.) is precisely that, evidence of something humans are

capable of *doing*, evidence of a type of activity, not in itself evidence that the processes underlying this activity are themselves representational.

4. If it is true that enactive accounts have not yet managed to explain the full spectrum of human cognition, it also remains true that computationalism (embodied or not) has not done so either.

Here we would like to gesture toward certain directions to further lay down our enactive, world-involving path, guided by some of these critical concerns. We will not describe a full enactive account of human action and perception (what this would even entail is not yet entirely clear), but we will show that, based on the ground covered in this book and the work of others, such an account is *feasible*, and we can even envision what it may look like in some of its aspects.

We start by addressing the already mentioned apparent limitation of the enactive account of action and perception: its inability to explain mental activity that is extended in time and scope beyond the current coupling with the environment, the so-called "representation-hungry" situations (Clark and Toribio 1994).

We will respond to arguments concerning the impossibility of enactive explanations of such cases. These arguments are usually based on a cocktail of (1) a misreading of the concept of sense-making and of the coupling between agent and environment, (2) the assumption that meaningful action relies on vehicles carrying content (i.e., applying a functionalist, not an enactivist, criterion for meaning) and that the dynamical coupling with the environment is an insufficiently versatile vehicle, and (3) an application of the "there-is-no-alternative" operator.

In the second part of the chapter, we propose a concept of virtual action based on the clustered network of schemes that constitute the sensorimotor agent introduced in Chapter 6. We suggest that the relations between schemes already activate a series of attitudes toward behavioral possibilities in addition to the ones that are enacted at the moment. This is done in accordance with the mastery of the lawful relations between schemes as well as the sensitivities that emerge from contextualization in activity genres (scheme subnetworks). The activation of these attitudes toward immediate and long-range behavioral possibilities opens the door for enactive explanations of more complex phenomena such as planning, rehearsing, and imagining.

Finally, we address the question of "perceptual attitudes," that is, the fact that human beings (at least) are able to switch between different modes of engagement with the world. From instrumental, goal-oriented attitudes to contemplative, abstract ones. The evidence in this case strongly supports the hypothesis that such attitude switching between instrumental and abstract forms of perception and action is enabled, and in some aspects even constituted, by social interactive experience and socio-linguistic skills. An enactive interpretation is sketched of how these enabling and constitutive relations work, thus suggesting the road for further work in world-involving, non-representational explanations of more abstract forms of cognition.

This ground will be covered as one sketches a rough map. It will provoke new questions, probably more questions than answers. But they will be interesting. The objective is to show that enactive explanations do not end in the embodied, complex dynamical stories of concrete agent-environment engagements. They do start there, in the concrete lived moment, and in a sense, they see all forms of cognitive activity, all forms of mind, as ever more sophisticated ways in which the concrete moment is *lived*. So they do also eventually come back to the concrete moment. It is possible that these explanations will eventually find their limitation. But we see no a priori reason not to continue the exploration of how far they can go.

8.2 **Modes of the mind**

In her book *Human Minds*, developmental psychologist Margaret Donaldson puzzles over the human capacity for constant invention of novel activities and goals. In some cases, these occupy a good part of our waking hours (obsessions with sports, hobbies, music, social networks, etc.). Some obsessions (the word still applies) involve high degrees of sophistication and demand high-level cognitive capacities (e.g., doing research, developing new technologies, understanding legal systems or financial markets) What is curious is that the study of human intelligence always focuses on how we learn to achieve ever more complex cognitive tasks, for instance, how we go from moving about without bumping into things to doing math, but hardly ever looks at the question of why we even bother to climb this complexity hill. There must be a motivation at the heart of every human action, but motivations will vary in kind. Of the different dimensions that we could use to describe the complexity of human activity and motivations, Donaldson proposes to look at a dimension of degrees of concreteness/abstraction in all aspects of the mind, including cognition, action, perception, and emotion. This is not the only interesting dimension, but it is quite useful in the current context.

This dimension is typically, and poorly, described as having two poles and many undefined cases in between. As the distinction between "higher" and "lower" forms of cognition can often be crude, embodied functionalists tend to speak of "offline" and "online" intelligence instead (Wheeler 2005). The first kind of intelligence is demonstrated by fluid embodied engagements with our surroundings (e.g., navigating a noisy train station full of moving objects and people without bumping into them). Offline intelligence, in one of Wheeler's examples, involves "mentally weighing up the pros and cons of moving into a new city" (p. 12). Such seemingly intuitive distinctions belie the inextricable Cartesian roots hiding in everyday language and the common sense from which they originate, the pervasive mediational epistemology Charles Taylor warned us about. The online/offline distinction mischaracterizes differences in cognitive complexity and reflexivity as levels of body involvement (the walking is done by the body; the "weighing of pros and cons" is done "mentally"). But I'm no less embodied and coupled to the world when I plan my holidays than when I ride a bike; I'm simply doing different things with my body and my coupling. To preconceive these differences in this way is

simply to use a different name for the separation between body and mind. Adopting the online/offline and related terminology means implicitly buying into a dualistic perspective. And uncritically assuming this dichotomy in the framing of the question is unlikely to lead to answers that do not perpetuate its inherent dualism.

Fortunately, Donaldson provides us with a more discerning alternative. She distinguishes four modes in which our minds function (for our current purposes, we simplify some important elements in her more elaborate proposal). Instead of drawing directly on a scale of sophistication, she proposes that these modes depend on our focus of concern and motivation. The first mode is the *point mode*. It deals with here-and-now coping. The focus of concern is the currently engaged activity involving a concrete ongoing motivation. Most animal activity falls within the point mode, and a large part of everyday human activity does as well. The point mode is the mode we engage when we make a pause while eating lunch and have a drink of water, when we avoid bumping into other people at the station, when we get dressed in the morning, tie our shoelaces, and so on, but also when we need to make a split second decision, anticipate a change in mood in our interlocutor, or estimate the rate of currency exchange while paying a cab fare in a foreign country.

Next is the *line mode*. It expands the focus of concern to the immediate past and the possible immediate future. For this reason, it may also involve a larger spatial region than the space I currently occupy and can easily reach. Picking up the umbrella on a cloudy morning before leaving the house is an example of line-mode activity. The line mode expands the point mode, but it often involves it at the same time: for instance, line-mode activity may concern things and events I perceive in the distance *now*, (I see someone I would prefer to avoid who is waving at me from the opposite side of the road) but whose consequences (the expected boring conversation) are *not yet* directly manifested. These potential consequences nevertheless allow me to regulate my activity now.

Next is the *construct mode*, which involves a decentering of cognitive activity. The concern of the agent focuses on events that have happened or may happen somewhere at some point in time. These events do not necessarily involve the agent herself; they could be about other people or a generalized other. Construct mode activity includes the recognition of different perspectives and points of view, generalizations about events, motivations, and behavior, and practical reasoning and norms (e.g., "it is best not to drink and drive"). A concrete engagement in the construct mode may still involve elements of the line and the point modes ("since it is best not to drink and drive, I will say no to the offer of another round of drinks that I can see my friend is about to make").

Finally, the *transcendent mode* involves no locus of concern as such. It involves concepts and relations between ideas, abstract contemplation, thought, values, and emotion.

As already suggested, cognitive activity is probably most of the time not performed neatly within a single mode. While it is reasonable to think that there is a developmental progression and that cognitive skills move from the point mode toward the transcendent mode with age and education, most of the time we are involved in some complex concrete mixture of activity in the different modes. For instance, one may be trying to think of the

best way to write a sentence that clearly encapsulates a complex idea, while at the same time being annoyed at some dog's loud barking and contemplating plans to move to a quieter neighborhood; all of this while exercising the skill of writing on an old laptop's keyboard.[1]

The idea of this classification is to have a relatively simple way of distinguishing between human activities that is nevertheless richer than the dichotomous, dualistic, and normatively biased opposition between offline and online or between high and low level mental activity.

8.3 The poverty of the coupling?

For embodied functionalism, in spite of important lessons learned from the embodied turn in cognitive science, representationalism remains the only game in town. We have already criticized this attitude, in part since it tends to be self-justificatory (define cognition as the manipulation of representations and there seems to be little alternative to representationalism) and in part because it is far from clear that the representational approach could produce a full, coherent and naturalized explanations of the mind.

But let us briefly consider what drives this kind of arguments, which often begin with an acknowledgment/dismissal combo ("yes, this is all fine; no, it doesn't change anything") of the lessons of enactive and dynamical approaches. As we have suggested earlier, the main drive behind these arguments is the supposed poverty of the coupling between agent and environment. If sense-making, as we have proposed, depends on adaptively regulating this dynamical coupling, it would seem that no sense-making could ever escape Donaldson's point mode. All sense-making, as defined by enactivists, would be limited to here-and-now sensorimotor engagements, at best.

It is true that the dynamical coupling between agent and environment plays a central role in sense-making. But why must this role be akin to that of a vehicle for carrying content? In fact, no one has ever made such a claim, or anything resembling it, except the embodied functionalists that build their arguments on its basis.

Embodied functionalists rely on the assumption that many cognitive problems are representation-hungry (Clark and Toribio 1994; Clark 1997), in other words, problems that deal with absent elements (not directly perceivable or realized) or abstract ones (including unspecified general goals). In Donaldson's categorization this would mean all mental activity in the construct and transcendent modes, and probably most activity in the line mode too. However, the idea that the class of such representation-hungry problems is defined univocally, or that, given a particular cognitive phenomenon, it is a straightforward matter to decide whether it is representation-hungry or not has come under considerable scrutiny and criticism (e.g., Chemero 2009; Di Paolo et. al 2010; Degenaar and Myin 2014).

[1] More examples showing how the different modes can be intertwined in everyday activities can be found in some of the few, but influential, studies that bring together cognitive science, ethnography, and anthropology (e.g., Agre 1988; Hutchins 1995; Lave 1988).

While Clark and Toribio (1994) have in mind what we could roughly describe as higher cognitive mental function when talking about representation-hungry problems, their proposed category would seem to include any kind of motor control problem for which direct sensorimotor feedback seems insufficient. For instance, performing smooth arm movements (or smooth movements in any other multijoint system) involves compensating for interaction torques that emerge on a given joint owing to the movement around another joint. Controlling for interaction torques would seem to be a representation-hungry problem because sensory feedback can be ambiguous (i.e., relevant information is absent). However, Buhrmann and Di Paolo (2014a) have demonstrated in a model implementing simple spinal circuit equilibrium-point control that the dynamics of the musculoskeletal system is able to compensate by itself for such interaction torques.

Similarly, Izquierdo and Di Paolo (2005) have provided a simple demonstration using a simulated agent of how some problems that could be deemed representation-hungry can be solved without the need (even without the possibility) of manipulating internal representations. They evolve the neural controller of a simulated robot able to move in one dimension using a simple array of distance sensors. The task is to approach or avoid a falling object according to its shape (a typical minimal cognition task in Randall Beer's style). However, without the agent having any way of knowing it, its sensor array may be reversed in the left-right direction. So if the falling object is to the right of the agent it stimulates the sensors on the right-hand side in the case of no reversion, or the sensors on the left-hand side in the case of reversion. In either case, the agent must approach the object. This task is a type-2 classification problem (one in which the instantaneous stimulus is ambiguous with respect to the correct moment to moment decision about the course of action to take, Clark and Thornton 1997). However, a robot controlled by a stateless machine can actually solve the task. Such a controller is incapable of handling any representation about the history of stimulation from which to extract relevant information to solve the task. Although from moment to moment, each instantaneous change in the movement of the simulated agent will be different depending on whether its sensors are reversed or not, the accumulated history of these changes still results in a trajectory that catches the object at the end. The key aspect of the non-representational solution is that the agent must be able to engage in a continuous coupling with the environment, and it is in this time-extended process that the ambiguities of the sensory input are resolved. This follows on from similar, and remarkable, demonstrations of simple embodied agents performing supposedly representation-hungry tasks, such as delayed, history-contingent choices, learning and conditional classification using the minimal cognition paradigm championed by Beer and colleagues (e.g., Beer 2003; Izquierdo et al. 2008; Williams et al. 2008).

In short, what is or is not a representation-hungry task is anything but a trivial or even a decidable matter. Perhaps the category should be dropped altogether (see also Degenaar and Myin 2014). The implication up to this point is that in explaining how we move beyond Donaldson's point mode into more sophisticated forms of action, non-representational explanations cannot yet be discarded. So far this seems a weak

clarification, but it gains weight in the light of the fact that this is precisely what many people do: they assume that there is no other way to solve some cognitive tasks other than through internal representations.

Consider the problem of "tracking" targets that are momentarily absent or of anticipating behavioral outcomes before the relevant environmental features are perceived (or even actualized). It is profitable to see how such cases can be explained in dynamical systems terms. The lesson we aim at is that whenever we consider an agent in coupling with the environment, it is not just the instantaneous coupling that matters for sense-making, but the history and dynamics of the agent and the environment as well. The very notion of anticipation has been considered from this dynamical perspective in the cases of ultrafast cognition mentioned in Chapters 2 and 7 (Wallot and Orden 2012). In these cases, fast cognitive performance (typically in go/no-go tasks) occurs reliably in times that are faster or of the same order as the estimated physiological limits of information transmission (to say nothing of processing). Since the explanation for this reason cannot involve any kind of anticipatory or predictive modeling (let alone more complex forms of representations), the problem must be approached differently. The agent is seen as a complex dynamical system, which is historically situated and able to move toward critical dynamical states. As the agent understands the possibilities of the go/no-go task, the coupled dynamics become "poised" in a metastable state, which can quite rapidly crystallize into the right cognitive performance. The stimulus acts as a trigger and not so much as bearing information that needs to be processed and decoded.

Evidence of ultrafast cognition supports the notion that cognitive systems can be strongly anticipative. According to Stepp and Turvey (2010), strong anticipation arises from the (ecologically situated) system itself via lawful regularities embedded in the system's ordinary mode of functioning in coupling with the environment. Weak anticipation, by contrast, arises from a model of the system via internal simulation. The distinction has been introduced in formal terms by Dubois (2003). A strongly anticipatory system is attuned to its situation by the integration of feedback, delays, and synergies into its overall dynamics. Before movement starts, the system is already in some form of attractor that will lead to the appropriate, context-sensitive behavior. Perturbations, like pulling on the chin while you are trying to articulate a phoneme (Kelso et al. 1984), simply result in taking a different transient toward the same attractor (equifinality with respect to the outcome, rather than to implementation of movement). In other words, the existence of the attractor means that the system already anticipates possible perturbations by allowing for different changes in its transients. And the setting up of the attractor is a sort of feed-forward action, hence by the time the movement actually starts, nothing more needs to be "computed," not even the appropriate response to perturbations; the agent-environment coupled system just relaxes toward a stable state.

This relaxation can indeed be context and history sensitive as demonstrated in Buhrmann and Di Paolo, (2014b), where an arm model with a touch sensor at the tip is able to follow a surface in front of it at various orientations and positions and then

continue moving along the same plane even when the sensor is turned off (as if predicting where the remainder of the surface "should be"). These trajectories, however, all lead as different transients toward the same attractor, and, in each case, it is the form of the transient that has been modulated by a history of coupling with the environment. Nothing in the system changes internally, nothing is represented.

The relative simplicity of these examples should not distract us from the important lessons they provide. The first lesson is that they show how complex action, perception, and cognition can emerge in an autonomous agent in coupling with the environment. There is no explanatory constraint from the enactive perspective to reduce the determinants of sense-making to anything less than the history-dependent coupled system as a whole. When we consider aspects of the coupled system and wish to understand their influence on the outcome of an action, we need always to make sure the question is properly contextualized. Meaningful activity, therefore, cannot be reduced to what may be found in a part of this whole coupled system, namely, the variables and parameters currently involved in sustaining the agent-environment coupling. These are *not* vehicles for content. The second lesson is that the agent is historically situated and able to change as a whole in terms of its history of interactions with the environment and other agents. It is precisely for these reasons that there is in general no such thing as "the function" of the coupling. The whole idea of a part or a process having a function is contingent on the current mode of stability of the system (in terms of the network of sensorimotor schemes introduced in Chapter 6, elements may or may not have a definite function depending on whether the agent is situated within an activity cluster or in transition between one cluster and another). Yet these historical changes in the agent are not required to take the form of representational traces, which are then available for inspection. Many other options are possible. Once we abandon the computer metaphor, we begin to realize that there are other ways of being affected by history and attuned to current and future situations. Yet a third lesson may be drawn, one that stresses the situatedness of the agent, one that may be easy to miss in the a-temporal economy of information processing. No matter how large the embedded scales of action, how far they stretch into the past or the future, or how far afield the decentered scope of cognitive activity is, in other words, no matter in which of Donaldson's modes action occurs, it always and *necessarily* occurs as the concrete acts of an embodied agent here and now. "The action itself is an event. It initiates novelty in the world. It is no longer a possibility moving toward the actual, but an aspect of the actual itself, the flesh of advancing process" (Ricoeur 1950/2007, p. 205).

If these lessons are true, then it must mean that sense-making is a very particular kind of activity, one that is grounded in the contemporaneous materiality of agency and the world and yet able to extend its sensitivities and its efficacies way beyond. And it is able to achieve this without the aid of representational traces or predictive models, but inherently in the sort of activity that it is (see also Haugeland 1995). To see this more clearly, let us consider the concept of sense-making again and, in the process, clarify some aspects about the relational ontology of enactivism.

8.4 Sense-making and virtual fields

As we have insisted, our account of mastery and sensorimotor agency is a world-involving one. It is an account where the world does not play the role assigned by functionalism in cognitive science: that of being a source of information. This sets the enactive approach apart from other traditions that have also stressed the importance of the relation between agent and environment, such as for instance the Gibsonian tradition in ecological psychology (Gibson 1979). There are many convergences between the ecological and enactive approaches. The ecological insistence on the richness of the environment as a source of information is a response to representationalists, for whom the purpose of perception is to do well within an economy (production, access, distribution, consumption) of a scarce commodity (information) which must be extracted, deduced, or even constructed by mental activity.

We agree with ecological psychologists when they highlight that real environments are rich enough to access directly their relevant meaningful aspects. We think they are in fact *too rich*, in that sense-making always involves a massive reduction of all the environmental energies that might affect the agent, to those within the dimensions of biological, sensorimotor, and social historically contingent meaning. However, the ecological response to the traditional view of the poverty of the environment is a move within the same conversation, one that does not question the nature of cognition as being in the business of managing information. They simply move from an economy of scarcity to one of abundance, but they do not switch discourse.

A world-involving account of a given cognitive phenomenon means precisely that the world is involved at more fundamental levels than as a source of information. In fact, it is one of the challenges of enactive cognitive science to provide clear theoretical reasons of what special circumstances allow us to speak of information processing as a limiting case, something that can happen, but under particular conditions.

What will happen, then, if we stop thinking of cognition as the management of content extracted from the environment that must be carried in vehicles from one place to another (see Hutto and Myin 2013)?

The first point to consider is that from an enactive perspective any form of action and any form of perception is a form of sense-making (Di Paolo 2005; Thompson 2007a). As we saw earlier, sense-making is the adaptive regulative activity of an autonomous and precarious agent. Through this activity, the agent is sensitive to tendencies in its states and its relations with the environment that would break through its boundary of viability. The agent responds to these negative tendencies by regulating their course so that such a breakdown may be avoided. No magic is involved in this idea as demonstrated by the various models of sense-making we have discussed earlier.

Let us expand on the notion of sensitivities and regulation of tendencies. The autonomous organism (the argument also applies to a sensorimotor agent) is capable of evaluating some interactions with the world as beneficial or not for its own continued self-individuation. In other words, the maintenance of systemic identity is a source of

intrinsic normativity. The problem is that the formal condition of operational closure of a living system is in itself binary, not graded. Some form of mediation is missing in this picture, something that can link actual states with the tendencies and capacities involved in evaluating the consequences for the system's viability. Since sense-making is the regulation of the coupling with the world based on this evaluation, it requires not only an autonomous organization but also an adaptive one as discussed in Chapter 5. Adaptivity is the capability of an autonomous system to respond to tendencies in the trajectories of its states and its relations to the world, so that when these tendencies approach the boundary of its own viability the system modulates its coupling with the world in a way that tends to avert the crossing of this boundary. Successful adaptive responses thus regulate states and relations before the breakdown of closure (death) has occurred.

Adaptivity, therefore, works on the *virtual field* that surrounds the current dynamical configuration of the agent–world system. In two senses: first, the agent must be responsive to whether or not tendencies approach the boundary of viability—if the crossing of this boundary is actualized, it is simply too late—second, it must make use of its capacities to modulate the coupling with the world by introducing changes that alter the virtual field in the vicinity of the current states (modifying the direction of the negative tendencies).

What exactly do we mean by a *virtual field*? The notion of the virtual in the current context follows from the tradition of granting real ontological status to capacities and tendencies, especially those in the concrete vicinity of current actuality. This tradition includes Aristotle's concept of potentiality, Spinoza's concept of affect, and related (but not identical) usages of the term "virtual" by Bergson and Deleuze. For our purposes, we additionally use the rich conceptual resources of dynamical systems theory to speak of trajectories, traces, and tendencies. In a few words, in this tradition the virtual is that which is real but not actualized, such as the glass's capacity to hold liquids or the tendency of liquid water to become solid at freezing temperatures. Such capacities and tendencies need not be currently actualized (this glass is empty and in principle could remain empty forever but its water-holding capacity is still real).

Notably, these capacities and tendencies are always *relational*, unlike properties of an object that belong to it in all contexts. Hence, virtual capacities and tendencies can be potentially infinite in number (the glass can be used as a paperweight; water also has a tendency to slow down high-energy neutrons inside a nuclear reactor). This does not mean that we can manipulate circumstances so that any object could acquire *any* arbitrary capacity or tendency (the glass does not have the tendency to slow down fast neutrons, and liquid water makes for a rather ineffective paperweight). Actualization of virtual capacities and tendencies occurs always as an *event* or a happening, or an act. These actualizations are situated in history, have duration, and so forth.

By speaking of *fields*, we want to focus attention on the concrete situatedness of virtual tendencies and capacities, that is, not just any abstract counterfactual, but those that neighbor a current state of affairs. This adjacent structured field can be studied by examining the dynamical landscape around a current trajectory (this landscape need not be

fixed in time). Such fields have been investigated in many simulation models of embodied cognition to different degrees of explicitness. It is common in such models to artificially travel back and forth along a recorded behavioral trajectory, manipulate circumstances, and rerun the model to ask "what-if" questions in order to map systematically a virtual field (e.g., Beer 2003; Iizuka and Di Paolo 2007; this is the method we have used to explore the sensorimotor habitat for our model in Chapter 3). The very concepts of stability and metastability in dynamical systems theory imply the idea of a virtual field since they point to structural aspects of adjacent flows that surround a currently actualized trajectory of states in a system. As we saw in Chapters 6 and 7, these ideas underlie complex concepts such as readiness-to-act and commitment.

We come to an intermediate conclusion: in a strict sense, all cases of sense-making, and therefore all action and perception, involve constitutively elements of virtuality.

In other words, sense-making, which even in its most basic forms entails adaptivity, always occurs in a "thick," extended here-and-now, not just in the actual changes of state of an agent but in its virtual field. By this we mean that, given the current situation, not only the actual states matter to the sense-maker but also the virtual traces and tendencies that surround these states, whether they become actualized or simply modified but not necessarily actualized in the course of events. Anchored in the here-and-now, sense-making always operates on a basis that transcends the present moment.

In the enactive analysis, making sense of a situation—evaluating its relevance to the sustaining of a precarious identity and responding accordingly—requires by definition both *sensitivity* to and *regulation* of the virtual field of possibilities. The sensitivity and capacity for regulation are present in all forms of life with large variations in complexity and refinement. For the average person, it is not the same to walk a straight course stepping along the curb at street level or stepping along the border of the terrace of a tall building, even if the sensorimotor trajectories that should be actualized are nearly identical. For the average pigeon, the two situations are similar. This is because both are differently sensitive to risks and have different capacities to respond to them. Apparently similar situations present to them radically different virtual fields.

We see then that, from an enactive perspective, normal action involves both sensitivity to the virtual adjacent field of possibilities and traces and a way of manipulating this field (for instance, by reducing risk or creating opportunities). From a sensorimotor standpoint, we have seen that this virtual field is modulated by the mastery of the laws of sensorimotor contingencies, which depends on the agent's history of adaptive equilibration, the structure of the sensorimotor relations between schemes, and its concrete situatedness within an activity.

8.5 **Virtual actions**

A useful concept that may help us shape future enactive explanations of how we move through Donaldson's four modes is that of virtual action. We take a virtual action to be an action that has not been fully actualized, and yet has real consequences for the agent and the world.

This idea inherits at least some of the explanatory burden of what in classical Artificial Intelligence goes under the name of planning. Whether in real robotic systems or in other contexts, planning in the traditional view involves maintaining and using an accurate enough representation of the world. Action-based approaches relax this requirement a little. For instance, Rick Grush, in his emulation theory (2004), suggests that internal models of the body and the environment, whose proposed primary purpose is to improve motor control, can be run "offline" to produce motor and visual imagery and explore simulated actions without actualizing them (see, also, Clark and Grush 1999).

Our proposal, predictably different, is that virtual actions involve the skill of extending sensitivities and regulations further along the virtual field and recursively invoke the virtual consequences of actions not yet taken.

This skill—for which we acknowledge that a more detailed operational account is not yet available in full—demands no new fundamental principle over and above the concept of sense-making. It does require, yes, more complexity. Consider how we have discussed the concept of mastery. This can be recast in terms of sensitivities to the virtual field conditioned on possible actions that are not immediately taken. We have mentioned in Chapter 7 the sense of being capable of actions and trusting the surrounding world, a sense of *I-can*, and speculated about its relation to the partial activation of several schemes in response to motivation and environmental solicitations. This is already a sort of virtual sensitivity. This sensitivity may become oriented toward a particular course. I see a cup in front of me and my perception involves constitutively my mastery of the virtual possibilities afforded by this cup and my body in interaction. I may move my head slightly or reach to grab the cup and my mastery of the relation between schemes in my sensorimotor network shapes my perception of the cup by the characteristics of these action possibilities, even those I do not take. We saw this in George Mead's and Margaret Floy Washburn's descriptions of the role of inhibition of viable actions in favor of a selected one (Chapter 2). To perceive in the sensorimotor approach is to master the influence of virtual actions (i.e., of possible adjacent actions, which are not taken).

What we need in order to explain more sophisticated forms of virtual action is an account that furthers this mastery into more complex sensitivities. Such mastery may only occur in organisms above a certain level of sensorimotor complexity. Such organisms act in the here-and-now but "with reference to a spread-out environment" linked to the spatially and temporally distant, "an environment both extensive and enduring [that] is immediately implicated in present behavior" (Dewey 1929, p. 279). As well as being a trace of recent events, the immediately present cup is like a "hub" for directly solicited virtual actions (move the head around it, grasp it, drink from it, etc.). A skillful sense-maker can regulate these immediate sensitivities to include virtual events projected further forward in time (grasp the cup now but then place it on top of another surface, near the desk at which I will sit to write a report).

There is no fundamental difficulty here, only the complexity involved in skillfully extending and navigating the virtual possibilities afforded by the here-and-now anchoring of sensorimotor activity. This complexity would be unmanageable if an agent were required to be

sensitive to all possible virtual actions. However, I rarely grab a cup to put it on top of my head or to fill it with sand. As we saw in our account of sensorimotor agency, I live within activities that cluster sensorimotor schemes into functional and structural self-sustaining relations. Within these activities, I live in a situated normativity that shapes the space of my current affordances (see, also, Rietveld 2008b, Rietveld and Kiverstein 2014). Inside a cluster, schemes "make sense" in relation to each other and this permits not only the regulation of immediately adjacent consequences but of whole streams of virtual activity at the moment I enact a single particular scheme.

We can illustrate what this would look like using our network description of sensorimotor agency.

Within a cluster of sensorimotor schemes, within an activity or behavioral genre, the enactment of a scheme, as we saw in Chapters 6 and 7, is influenced by previously enacted schemes and by those schemes that are currently inhibited or activated to be enacted in parallel. The enactment of a scheme has virtual implications for the future propagation of the activity. These past, present, and future influences shape the enactment of the scheme here and now, and the enactment in turn modifies these virtual influences. I grasp the cup differently depending on whether I see tea in it and want to drink some or whether I will wash it in the kitchen sink or whether I will put it back in the cupboard. A skilled agent may be more sensitive to such influences, which may then involve activity further back or forward in time or space. I grab the cup differently whether, for example, (1) I aim to put it in the cupboard, close the cupboard door, turn around, and leave the kitchen or (2) I aim to empty the cold tea, rinse it, put the kettle on, choose a new tea bag, and wait till the water boils. The differences may be only slightly manifested in the actual movements and in postures that put me on course to the next events in my activity stream. In this way an agent regulates the relations between schemes within a cluster of activity. Eventually, these regulations may not take the form of an overt, enacted scheme that has been prepared by sensitivity and readiness to ever-longer series of projected actions. To act virtually is like any embodied skill, and practice may lead to progressive refinements. At some point the virtual act may simply be the act of regulating long-range sensitivity and readiness, even before acting, by a prolongation of the "onset" phase of an enactment. Most of the time, however, these regulations happen while we act. We are thus frequently *thinking in movement*, to use Maxine Sheets-Johnstone's phrase (Sheets-Johnstone 1981).

There may be different processes involved in these regulations and sensitivities, which count as virtual actions; some of them were discussed in Chapters 6 and 7. In Figure 8.1 we illustrate a hypothetical situation. On the left we see a part of a sensorimotor network. A scheme in the blue cluster (S_1) is currently being enacted. This facilitates the subsequent activation in the agent of the support structures for schemes S_2 and S_3. We assume that the environmental conditions allow either of these two to be enacted next. But S_2 and S_3 tend to inhibit each other and cannot be enacted together. In addition, each of them leads to the facilitation of schemes in different clusters (e.g., the activity of preparing a new cup of tea or the activity of leaving the kitchen after putting the cup away). Even before the competition between S_2 and S_3 is resolved, the activity in the support structures for

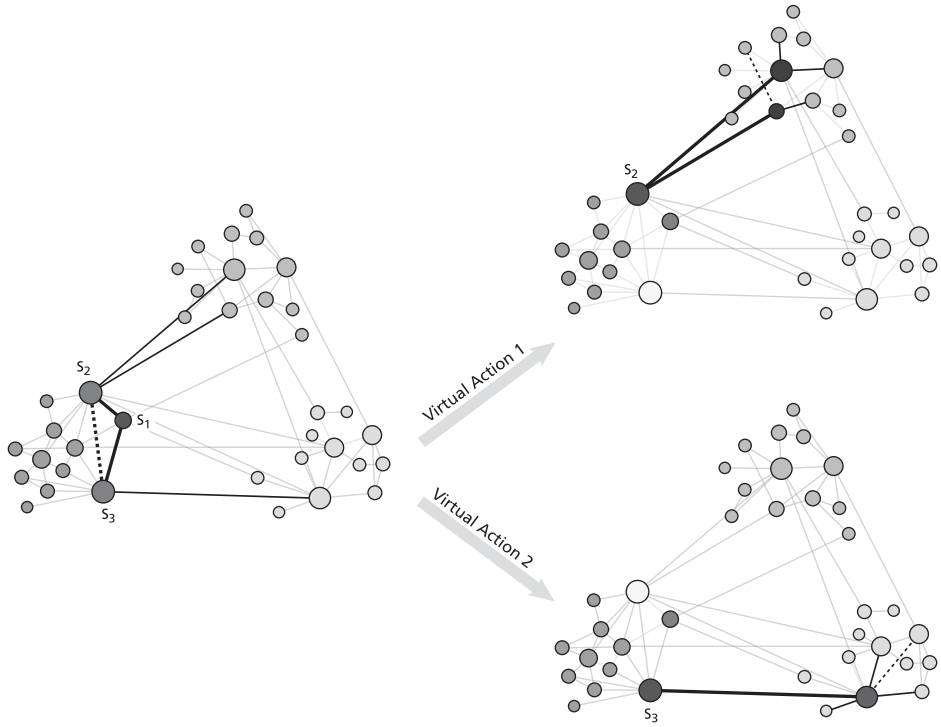

Figure 8.1 Hypothetical section of a functional network in a sensorimotor agent. **Left:** A single scheme (S_1) is currently enacted, as indicated by its black shading. At the same time, through the influence of S_1 as well as the potential contribution of environmental solicitations, two other schemes are already "primed" (S_2 and S_3, darker shade). This could involve, e.g., the "subthreshold" activation of the support structures in the agent required for these schemes. Though S_2 and S_3 have no overt effect yet, their latent activity may nevertheless already influence yet other schemes in the network (illustrated as thicker connections to nodes in different clusters), which may as a result become more likely to be enacted in the future. These "primed" schemes constitute a strongly anticipative context for the current action, which thereby extends its horizon of influence into the future. The two schemes S_2 and S_3 are also mutually inhibiting and thus they compete with each other for enactment. **Right:** activity shifts to either S_2 or S_3 and one of the two schemes is enacted (top and bottom). Several events in the agent can cause the tipping of the balance between the two competing schemes. This may be the influence of other active nodes in the network on S_2 or S_3 (not shown), an uneven influence of S_1 on the two schemes, or a more holistic act, such as the agent changing the relation to the environment with the result that one affordance becomes more salient than another. All these are examples of virtual actions, and their result is the shifting of activity in relevant parts of the network. Note that the agent-side structures involved in previously enacted schemes may continue to exhibit influence on other parts of the network, even when no longer having any overt effects. Such "lingering" activity extends a currently enacted scheme's horizon of influence into the past, which current activity can resignify. (See color plate.)

Copyright © 2017 Ezequiel A. Di Paolo, Thomas Buhrmann and Xabier E. Barandiaran, with permission.

these schemes at "subthreshold level" may already pre-activate schemes in the green and yellow clusters. The situation is that of a system poised at a critical state. Moreover, the influence of S_1 and the way it is enacted (which is not entirely up to the agent since the environment is involved) may still "linger" in its support structures (e.g., neuromuscular processes, posture, inertia). Maybe the influence of other recently enacted schemes also lingers. A virtual action in this case may take the form of tipping the balance between S_2 and S_3. It may occur by the way S_1 is enacted in actuality, which already orients the flow of activity toward one of the options, or it may occur as an emergent down-regulation of the agent as a whole (e.g., through activity in other parts of the network that influence either S_2 or S_3). The result is that the functional network takes a different shape (Figure 8.1, right) depending on this "internal gesture." Note that the spreading of latent activity in the network, that is, the influence of schemes enacted in the past and the influence of currently enacted schemes on those that may be enacted in the future, not only biases *which* schemes are activated and when, but may also modulate the way in which they are enacted. This is what we mean when we talk about currently enacted schemes being "colored" by previous ones and in turn "coloring" future ones—or what we have referred to as the "thick" here-and-now.

This is one possible example showing how the agent may regulate a virtual flow, in this case through changes in the relation between the agent-side support structures for interlocked schemes. It is important to note, however, that the example should not be interpreted as suggesting that virtual actions are somehow "internal," which would take us close to a representational story. Firstly, they only count as actions because they are the regulations induced by a whole agent, even if what in some cases is regulated may be the relations between internal support structures. Secondly, virtual actions involve changes to the relations between schemes without necessarily activating all of the schemes involved; this is why we call them *virtual*. But they are *actions* because they have consequences for the relation between agent and environment, more precisely adaptive consequences for the ongoing individuation of a way of life. In this sense, virtual actions are like other forms of sense-making, an activity that must be enacted, and that may fail or succeed.

In addition to future oriented regulation of activity flows (what shall I do next and how?), virtual actions can also help us fill in gaps and overcome micro-breakdowns with an activity. I see a meaningful course of action solicited/enabled by the environment, but the activation of the most suitable sensorimotor scheme to embark on that course is not possible in the current situation (say, I cannot find the pen I need to take down a phone number, or the hand I normally use for writing is injured). Hence, I must fill in these gaps with schemes that enable the activity I aim to enact. The enabling relations between schemes guide this search for alternatives, which again is performed by the agent as an emergent regulation of how schemes relate to each other, in other words, as a virtual action.

It is important to stress that all action—all sense-making—contains elements of virtuality. Virtual actions are, in particular, actions that regulate virtual fields, that is, the

relations between networked schemes within clusters of situated activity.[2] They take place in the shaping of activity flows before they are actualized, as well as in the reshaping of the virtual implications of *past* actions, which are then re-signified by current enactments as the agent orients her activity toward future possibilities.[3] In the sensorimotor realm, anticipation and memory can be aspects of one and the same thing.

Whether in cases of future orientation, or in the reshaping of the meaning of past virtualities, or while she finds alternative routes toward a goal in the form of sensorimotor planning, rehearsing or imagining, the sensorimotor agent is not required, under this view, to represent the relations between schemes in an activity (as *we* represent them in our network diagrams) to achieve any of these things. Since these relations are the same relations that constitute her as an agent, every reaffirmation that she enacts and every breakdown that she suffers and overcomes contributes toward developing her sensitivities to *herself* as an agent, and eventually, toward maturing a practical, embodied form of self-mastery. We see here yet another meaning of the concept of sensorimotor mastery, this time applied to the contingent "laws" and norms embodied in the precarious organization of a sensorimotor agent as a result of her personal history. This meaning makes explicit a connotation of mastery as *power* that is only implicit in the other meanings we examined before. The emerging and reflexive power and self-understanding of the sensorimotor body becomes systematically potentiated during development with the acquisition of language and, with it, of socially enabled modes of self-regulation and self-control (Cuffari, Di Paolo, and De Jaegher 2015).

A consequence of our view is that only a *sensorimotor* agent can be capable of virtual actions of this kind. A biological agent that has not achieved sensorimotor autonomy would then be unable to perform this kind of manipulation of the virtual relations between schemes because it lacks both (a) the normativity that is established by the networked individuation of precarious schemes, and (b) the capacity to modulate its scheme network activity in relation to these norms. This does not mean that all sensorimotor agents would necessarily be capable of virtual actions, only that virtual actions "occur" in the regulation of the various structural and functional influences subtended by schemes in a closed, self-sustaining network.

[2] We note that our proposal for a non-representational account of virtual actions is highly compatible with ecological approaches. Indeed the notion of affordance is a paradigmatic case of a virtual relation between organism and environment, which should be considered in its activity-dependent situatedness (Rietveld and Kiverstein 2014). A recent proposal by Brincker (2014) elaborates a similar perspective using the notion of affordance navigation to give an account of complex cognitive phenomena such as the passing of false-belief tests.

[3] The idea of current action reshaping the past is explored in the context of social interactions (Di Paolo 2015). When engaged in them it may be possible to own personal intentions retroactively once an action is completed socially (e.g., through the response of others). As noticed by Elias (2015) and Wallot (2015), this retroactive effect, here understood as a present reshaping of the virtual implications of past events, need not only occur in cases of social interaction but can happen more generally.

8.6 **The zero mode**

We noted in Chapter 6 that we sometimes find ourselves at the hinges between activities, those moments where we must begin to enact a new microworld. These transitions may be more or less noticeable, more or less dramatic, but they are frequent. They sometimes imply some kind of breakdown, as changes from one activity to the next do not always occur in a smooth fashion (activities are often interrupted, put on hold, changed on the fly, etc.)

These moments of breakdown are among the most mysterious and less studied aspects of the human mind. These are precisely the moments where functionalism itself breaks down (hence the prevailing focus on task-driven studies in psychology and neuroscience, with the exception of more recent studies of default brain activity and studies involving meditation). At these breakdown points, the coordinates that grant stability to functional descriptions are in flux. They are a blind spot for functionalist theories. However, although we may be momentarily world-less in a sense, we are not mindless. In fact, in coping with these moments our autonomy as organisms and as agents is reasserted. So it is indeed a way of being a human mind to be lost and in search for new coordinates, and eventually to find them. But this way of the mind does not fit precisely in any of Donaldson's four modes, since they are all defined in terms of the locus or focus of concern. What occurs during transitional breakdowns is that concern is precisely *unfocused*. Goals may be confused, actions misdirected, aims fruitlessly multiplied and abandoned without properly starting on a new activity course. We call this the *zero mode*.

According to Varela (1992), this is where enaction occurs, precisely in navigating oneself outside this nebulous state back into a microworld. Varela conceived these transitions as relatively fast (fractions of a second; in our network model this would also be the case as one needs just a few jumps to get from one cluster to another, while one can remain jumping between schemes within a cluster for longer periods). But in some cases, the reconstitution of meaningful concern can take longer. For instance, if we are engaged in an activity and someone comes to tell us some shockingly bad news, some of us may drop what we are doing and attend to the new and unexpected implications of this event. But others may find it difficult to disengage from what they are doing, partly because the shock may not have fully registered or partly because they cannot find what is the appropriate reaction. So for a while their previous activity continues in an uncanny manner.

Another situation in which the zero mode is persistent occurs when sensorimotor skills are systematically disrupted, as we have already seen in Kohler's visual inversion experiments. In such cases the clustered linkages between schemes are thrown into disarray and must be reconstituted almost entirely. The subject who is starting to wear visual inversion goggles is unable to engage fully in normal activities. For some time, she will always find herself back at the starting point of getting the first scheme right. And it is precisely how this scheme is conditioned by what should follow next that makes the adaptation process so difficult, because what should follow next no longer makes the same sense as before the visual inversion. Thus in order to grab a cup to place it in the cupboard the untrained

subject needs to overcome first the overall aim of her arm and hand movements, but even when this is overcome the scheme will not be ready for what comes next as the cupboard may be perceived in the wrong spatial position. With training not only are schemes enacted with some success, but the relations between them begin to be reconstituted as well. During this period the zero mode of activity makes itself very apparent.

In the same way that concrete activity may involve more than one mode, elements of the zero mode may also be present, for instance, in a creative resolution to micro-breakdowns (breakdowns within an activity that can normally be recovered by the adaptive relation between schemes). This could lead to a reconfiguration of the network of schemes, even to the appearance of totally new ones. Activity in the zero mode is what drives spontaneous creativity in agents capable of sensorimotor open-endedness (as described in Chapters 4 and 6). We can expect then the zero mode to be active in situations of play as new elements, rules, people and situations are added to a game, and the self-defining purposes of playful engagements shift. Successfully traversing the zero mode is creative sense-making at its purest. However, this only occurs as the result of the precariousness of the material relations that constitute the agent and the environment. It is because of material precariousness that breakdowns will occur and adaptive activity must restlessly reconstitute meaning by finding states and relations that are more viable if the organization is to survive. And it is because of the materiality involved that this reconstitution of meaning can in principle become open-ended.

8.7 **Abstraction and perceptual attitudes: the missing social dimension**

Virtual actions as we have defined them, the active regulation of virtual fields that affect sense-making, are the counterexample to claims that non-representational approaches are incapable of explaining cognition across multiple scales, and beyond Donaldson's point mode. But they cannot be the whole story. We still need to understand how it is possible to train these extended sensitivities, and to adopt more reflective attitudes on the activities we engage in.

At various occasions throughout the preceding chapters, we have remarked on the need to include a social dimension in our enactive approach to action and perception. A proper treatment of the complexities brought into our proposals once we consider that sensorimotor agents are also, in general, social agents, is beyond the scope of this book. The interactions between biological and sensorimotor agency are complex and hard to untangle. Adding a social dimension to these interactions, and, as argued by Cuffari, Di Paolo, and De Jaegher (2015), including properly social forms of agency into the picture, will apparently make matters even more complex. This may be so. Then again, it may be the case that, as it often happens when dealing with multiple levels of analysis, several unresolved issues in one dimension can only be resolved once another dimension is taken into consideration. Be that as it may, in this section we will limit ourselves only to exploring one particular issue that is directly relevant to the current discussion regarding the

feasibility of the enactive program: the question of how an agent is capable of adopting abstract perceptual attitudes. This again will show that the enactive approach can offer explanations for phenomena beyond Donaldson's point and line modes.

8.7.1 The abstract object, a false starting point

We said earlier that sensorimotor agents are generally involved in some concrete meaningful activity, a microworld. In general, this is implicitly acknowledged in embodied theories of perception. Yet, once we consider the starting points for most of these theories, they do not often involve the agent embedded in some concrete activity, but instead they consider a relatively demotivated agent confronted in the abstract with some external object. Most of the examples that illustrate the sensorimotor approach to perceptual experience are of this kind (see Chapters 3 and 4).

Like in representational theories (Marr 1982), it may seem that the simplest starting point for theorizing about perception is a situation involving a single agent confronted with a single static and neutral object. More elaborate cases incorporating multiple objects, in motion, available for different uses, or set within naturalistic scenes and social situations, must follow this prototypical case as so many degrees of increasing complexity. Everyday experience and experimental evidence tell a different story. I am hardly ever confronted with abstract objects, which I regard and inspect in their perceptual presence in a manner devoid of interest or social significance. Most of the time, I work my way around a world of social and biological purposes and norms, for which objects are seldom present to me as the bottles or tomatoes that populate the sensorimotor literature (O'Regan and Noë 2001, p. 945; Noë 2004, p. 62). I don't normally inspect a bottle with my fingers to perceive in it a Gestalt shape as opposed to disjointed sensations; I pour some beer out of it. I don't regard the red tomato and consider its unseen sides and its voluminous presence; I cut it and add it to the salad. The problem of how the object maintains its identity despite its manifold of appearances is not a problem for an agent whose attitude is pragmatic, but only for one whose attitude is abstract. Once I pick up a hammer to drive a nail, it is clear to me that it remains the same hammer between blows, as its identity is verified by its being part of the sensorimotor schemes that I enact.

Given our account of how sensorimotor agency is constituted as a self-sustaining network of schemes organized into clusters of activities, a pragmatic, instrumental attitude in the sensorimotor agent is always more primary that an abstract one.[4] But then we should ask how could the latter emerge from the former. How does the sensorimotor agent ever develop such a stance in regards to objects closer to Donaldson's construct mode (exploring the shape of the bottle or regarding the tomato as a soft red volume)?[5]

[4] We must not forget that the pragmatic attitude is itself a developmental outcome. Hence it is more primary than the abstract attitude, but not altogether primordial.

[5] While framed as a problem of abstract perception of objects, the same questions apply to the capability of producing abstract acts. In his classical studies of abstract and concrete behavior in patients

It has been suggested that the answer to this question must implicate the social world (Froese and Di Paolo 2009; Gallagher 2008; 2009; 2014). Shaun Gallagher aptly describes embodied theories of perception, and more traditional ones, as *philosophically autistic*. Their guiding assumption is that of a transparent, literal, and accessible objective reality that contrasts with an opaque, complex, and half-hidden social world. This assumption applies pretty much across the board in the sensorimotor approach literature. It is implicit in the unjustified supposition that one must first investigate how we perceive objects and only then move on to investigate how we perceive the social world (e.g., Engel et al. 2013).

Were it not for my co-inhabiting a world in which a myriad of perspectives, motivations, and interests flourish, I would never learn that an object has a presence beyond my own "solipsistic" concerns. It is because I can somehow see that I am not the only one for whom the object is meaningful, that I can also appreciate as an abstract object what remains at the intersection of those motivated perspectives—which "cancel out" like the summation of so many randomly oriented vectors.

Our approach to sensorimotor agency has filled in a gap in sensorimotor contingencies theory and addressed the criticisms by Thompson (2005) in a non-representational manner (as advocated by Hutto 2005). But our proposal leaves us with a problem lying, so to speak, at the opposite end: if perception is so constituted by *my* enactment of sensorimotor schemes, *my* deployment of *my* skillful mastery of the laws of the sensorimotor contingencies that correspond to *my* sensorimotor body, why is it that I perceive a world that transcends my activity? Why, in other words, do I perceive objects as being out there, publicly present to me and to others?

Even after attempting to resolve this issue through a story describing the complexification of a sensorimotor network through differentiation, integration, and clustering of schemes (Chapter 6), we will still find the problem unresolved. At the end of such an account, the agent would access the external object in multiple ways: the confluences of several modes of activity-dependent engagement between body and object. But is this sufficient for the object to have an existence that transcends the agent's concrete activity? The object has its own synergy with the agent's body and so it is manifested in a particular way, one that belongs to this object and not to another one. It is a synergy of the agent's modes of seizing the world, a set of organized relations of sensorimotor schemes. Two objects may be seen as different, but the agent sees both as emerging from her own activity. The object, distinct from other objects and in relation to them, remains

with neural lesions, Kurt Goldstein correlates the lack of a capability in dealing with abstraction (e.g., arranging objects in arbitrary geometrical relations and performance of meaningless movements) with impaired capacity in patients to put themselves in the place of another (Goldstein 1963 p. 55, based on lectures delivered in 1940). This is in line with what we will suggest in the rest of this chapter. The link between lack of understanding of abstract attitudes and lack of understanding of other perspectives is suggestive of a common capacity for decentering which is impaired in these patients. It also helps explain their socially awkward or misplaced emotional reactions as whole situations are understood only partially through disjointed concrete elements.

inescapably oriented toward the agent's interests and norms if we were to apply the equilibration story developed in Chapter 4. It could not be otherwise. What distinguishes an object is the achievement of the multiple equilibration events of the agent's sensorimotor schemes, a process whose norms are those of sensorimotor agency. Perceived objects therefore achieve only a relative independence; independence from each other, but not from acts regulated by biological and sensorimotor norms. A sensorimotor agent of this kind inhabits a "solipsistic" world.[6]

One solution to this problem, as indicated, is to go social, to include perceptual acts whose normativity transcends that of the constitution of precarious sensorimotor agency. Let us examine some of the evidence that support this move (see Di Paolo [2016] for further details).

8.7.2 Social determinants of perception

In a broad sense, perception is influenced by the social world because objects themselves, the situations they are embedded in, and the practices they involve are largely social in origin (Costall 1995; Heft 2007). Infants play with socially produced toys and wear socially produced clothes. To this we should add the social origins of various sensory standards (musical scales, prevalence of certain geometrical shapes and color palettes, etc.). This is also the case even for "natural" objects encountered by the developing child: they are subject to socially established norms and practices (you don't take insects to your mouth, you don't yank out the pretty flowers, etc.).

This is not a trivial observation. It points to the possibility of cultural factors affecting perception. These factors are indeed found to co-vary with perceptual "styles," even if we restrict ourselves only to the visual perception of object and scenes. Cultural influences are found in the perception of visual illusions (Segall, Campbell, and Herskovits 1963), perception of depth and orientation, and pictorial representations (Deregowski 1989). A well-researched cultural variation in visual perception is the contrast between holistic attention to context in East Asian cultures and analytic attention to parts in some Western cultures (Nisbett and Miyamoto 2005). Holistic viewers show greater skill in attending to

[6] The problem is an old one. Phenomenologically, the issue of object presence is closely related to the problem of the transcendence of the world with respect to subjective experience. The argument regarding the insufficiency of a single developing subject undergoing progressive equilibration of her sensorimotor organization for constituting an abstract, objective perceptual attitude parallels Husserl's argument against the solipsistic constitution of the objective world through the variations between normal and pathological or disturbed experiences (Husserl 1989; see, also, Gallagher 2008). Like Mead (1934, 1938) who viewed the meaning and objectivity of the world as a consequence of the socially acquired expertise at adopting various roles, the Husserl of *Ideen II* sees the transcendent as constitutively intersubjective, as does Merleau-Ponty (1964). Each of these philosophers (despite important differences) seems to appeal to a similar solution involving the transformative consequences for the subjective grasp on the world that depend on effectively coordinating perspectives with others.

background and relations between objects while analytic viewers excel at discriminating figures from backgrounds (Witkin and Berry 1975). Such differences in visual style are manifested from a certain age onwards, but not before (Duffy et al. 2009).

Of course, cultural variations in perceptual styles do not immediately imply an interpersonal influence on development. It is also possible that other factors are at play, for instance, the structure of typical visual scenes (Miyamoto, Nisbett, and Masuda 2006). Other relevant factors tend to co-vary with cultural ones (climate, diet, landscapes, etc.) and these could have effects on perceptual development. While they are suggestive of the influence of social practices, cultural variations are not definitive proof. We can at least claim that they play a contextual role in modulating perceptual styles, i.e., they affect the style of perception but do not seem to enable its possibility (see De Jaegher, Di Paolo, and Gallagher 2010). Conversely, lack of cultural variation is not by itself sufficient to discard social influences on perceptual development.

Social cues also affect visual attention and object salience. The gaze of others influences affective judgments with participants finding objects more appealing if they are congruent with perceived gaze direction (Bayliss et al. 2007; Hayes et al. 2008). Eye-tracking and event-related potential measurements indicate that 4-month-old infants are sensitive to the difference between directed and averted gaze with respect to an object (Hoehl et al. 2008) and can be similarly sensitive to other cues such as head movements. Again, such influences modulate object perception, but the evidence so far does not prove that these influences play a strong enabling role in the development of perceptual skills, only a contextual one.

Evidence of stronger social influences on attention and manipulation capabilities in infants can be found in the development of triadic competencies during the first year of age. During this period infants become increasingly sensitive to joint-attention contexts as they move from dyadic interactions with the caregiver to triadic interactions also involving objects (Striano and Stahl 2005). Primary intersubjectivity (i.e., early reciprocal engagements, affective tuning, and mutual regulations between caregiver and infant) is progressively followed by co-regulated engagements involving shared attention and action on objects, or secondary intersubjectivity (Trevarthen and Hubley 1978). Dyadic interactions are established early on. Infants attempt to re-engage interactions that are paused artificially (still-face paradigm) already in the second month (Tronick et al. 1978). At around 6 months, objects dominate the infants' attention while they sometimes look up at their mother. Between 9 and 12 months, they engage in full triadic exploration of objects paying attention to both the adult and the object and focusing on the relation between the two.

There are close developmental links between dyadic and triadic competences (Striano and Reid 2006). This is confirmed in naturalistic longitudinal studies. Kaya de Barbaro and colleagues observe robust changes in the infant's sensorimotor organization that go together with changes in the organization of social engagement (de Barbaro, Johnson, and Deák 2013). At 4 months infants engage all the sensory modalities in the exploration of the object in front of them, grasping it, looking at it, taking it to the mouth (see,

also, Smith, Yu, and Pereira 2011). This is followed in the second half of the first year by more frequent separation of activities allowing decoupled attentional engagement with the adult and other objects in synchrony with object manipulation. Mothers are active during this transition in engaging the infant's attention to the object (showing it, moving it in conjunction with smiles and vocalizations). They solicit eye contact and attempt to switch the infant's attention from one object to another. These solicitations result in the infant at 6 months being able to decouple attention to objects and parent through different modalities, holding one object in the hand while looking at another one or attending to the mother. Transitions between objects or holding two objects at once go together with a reduction in the mother's bids for attention to novel objects and more engagement in attending together to the same object, imitating, and facilitating the activity of the infant. At 12 months the infant displays bi-manual coordination in handling objects in elaborate sensorimotor schemes. Mothers at this stage perform manipulations of objects that are more complex, and the infant attends to these demonstrations and alters subsequent manipulations. At this age, interaction patterns organize in activity turns and involve partial acts such as giving or requesting.

These and other naturalistic observations (e.g., Nomikou, Rohlfing, and Szufnarowska 2013; Rossmanith et al. 2014) suggest that the infant's attention and sensorimotor skills are "tutored" by context-sensitive scaffolding resulting in a socially guided mastery of attention and object manipulation. These social influences are more than merely contextual, they are enabling in the sense that these experiences play a role in shaping a particular kind of sensorimotor network in the infant.

As the child grows, she progressively adopts decentered attitudes toward objects and others.[7] Interpersonal coordination, breakdowns, and recoveries of coordination that make up a social interaction (De Jaegher and Di Paolo 2007) seem to play an enabling role in learning to decenter. In the 1970s, a team of psychologists at the University of Geneva studied the effects of interactions while children attempted to solve a problem together (e.g., Doise, Mugny, and Perret-Clermont 1975; Doise and Mugny 1984). The problems were typical Piagetian tasks: conservation of length and volume, rearranging objects to a new location conserving their relative positions, as well as games that could be played alone or in collaboration.

[7] Adults are much better at decentering than children, but not infallible. They seem to achieve a decentered solution to a problem only after having considered and inhibited an egocentric one. For instance, adults instructed to move an object in a shared frame of reference often look at the egocentric solution first (a salient object they can see but is occluded for the instructor) and so do children (Epley, Morewedge, and Keysar 2004). However, adults inhibit this initial reaction and proceed to reach toward the correct object visible to all while children have more difficulty doing this. This is compatible with the notion of decentering involving an active regulation of one's own perspective from the perspective of a real or putative other, rather than an acquired and readily available novel kind of personal "objective" perspective, in which case one would not expect adult egocentric looking to occur at all. Decentering, it would seem, is a virtual *act*.

In these experiments, children are individually pre-tested on the task in question and ranked as low, intermediate, and high performers. Then they are allocated into pairs to solve a similar problem together, and finally tested on their own once more. Provided that certain conditions apply, there are reliable positive improvements in individual performance after the interaction. Doise and colleagues observe that individual progress depends on interpersonal conflicts during the joint condition and on how they are resolved. Improvements in performance do not correlate with the observation of the correct solution to a problem, as for instance when pairing a passive low with a dominant high performer. They correlate instead with the degree of participation of both children in working out the task together. These experiments show a clear enabling influence between interactive factors (conflict, participation, breakdowns, and agreement) and progression in sensorimotor/cognitive tasks that require a decentered attitude.

Finally, we should also consider the influence of language and play in shaping, enabling, and possibly also constituting the capability to adopt abstract perceptual attitudes. These effects are major and not fully understood. For Vygotsky (1978), linguistic mediation is the key factor that differentiates human from animal perception. Lexical categories structure the perceptual world (e.g., Brown and Lenneberg 1954; Luria 1976; Rosch et al. 1976). As we have suggested earlier, linguistic abilities endow an agent not only with tools for understanding the world but also with instruments of self-mastery, for regulation of virtual fields. Language enables the control of perception by permitting the voluntary shifting of attention, easier retention of past perceptual events, and action planning. These influences are too numerous to describe here in detail. Even if we limit ourselves to the non-overtly social aspect of perception, evidence shows that there is a deep influence of language on color perception (Thierry et al. 2009), object recognition (Boutonnet and Lupyan 2015), motion detection (Meteyard et al. 2007), and spatial understanding (Richardson et al. 2003).

Flexible relations to objects during mother–infant interaction underpin object substitutions in make-believe play, which are at the basis of symbolic function (Vygotsky 1978). Meaning substitutions work when the child adopts the right embodied gestures (a pencil becomes a rocket ship by re-enacting a countdown to lift-off and imitating the corresponding movements and soundtrack). This capacity becomes progressively abstract during the second year (Watson and Jackowitz 1984) allowing substitutions to depend less on similarities (a toy banana for a phone) and increasingly on linguistic/gestural schemes, usually in a context of social interaction. As single make-believe episodes turn into social dramatic play (Starks 1960; Bodrova and Leong 1998), children learn to not only self-regulate and adopt different social roles but also to perceive and use objects as the recipients of novel social meanings, no longer just centered on the immediate sensorimotor grasp (Sinha 2009).

The developmental relations between language-mediated play, language comprehension and use, the development of malleable object relations, and the self-regulation of attitudes and roles are not fully understood. But it would seem that several of these processes are of the essence for the capability to regulate perceptual attitudes. To be able to adopt a

detached observational attitude toward an object, the child must be able to empty it of its most immediate meaning, to temporarily bracket its significance. This capability logically, not just empirically, implies the power to manipulate meanings and relations between objects, uses, and perspectives and this suggests that some linguistic and interactive abilities play a constitutive role in abstract object perception.

8.7.3 Abstraction, an enactive interpretation

From an enactive perspective, any explanation of the emergence of decentered perceptual attitudes must be based on how the agent's sensorimotor organization varies through processes of (metastable, precarious) equilibration. Dynamically, such processes rely on the emergence and subsequent resolution of conflicts: conflicts within a given scheme (adjusting the strength and precision of the grip upon realizing that the cup is a bit too full), conflicts between schemes (adjusting to driving on the opposite side of the road), and interpersonal conflicts. The latter seem to be at the root of the most compelling evidence briefly described in the previous section. We see this in the experiments by Doise and colleagues, where interpersonal conflicts that afford some degree of joint participatory resolution are the ones that most influence individual progress.

Conflict is also present implicitly at the phases of qualitative change in mother–infant dyadic and triadic interactions. In the observations by de Barbaro and colleagues, we see mothers insisting on contextually diversifying the infant's acts and attention when these are multimodally over-focused on a single object and later encouraging refocused attention and engagements as the infant learns to manipulate different objects at the same time. It is as if mothers were providing a context aimed first to enlarge the infant's set of sensorimotor schemes, let them accommodate to each other, only to follow this with a context aimed to refocus the now diverse modes of engagement so as to make them work together. This guidance is conflictive (though this is not to be understood as antagonistic) since the mother's influence tends to dis-equilibrate the infant in specific ways.

If we represented these transitions using the network description of sensorimotor agency, the stages would look as in Figure 8.2. A first stage (at around 4 months or less) shows highly synergistic schemes. The infant focuses on the object with full attention and multiple modalities. As the mother induces disequilibrium, these schemes begin to differentiate. At around 6 months, the infant is capable of manipulating and attending to more than one object and to the mother. Schemes begin to be grouped by different activities. Now the mother induces another disequilibrium by soliciting that the infant be more attentive to specific objects. This attention is shared and the object is manipulated together. The diversified schemes begin to accommodate to each other and form new links between them toward the end of the first year as shown in the figure.

These examples of *interpersonal* equilibration, as well as those described by Doise and colleagues during peer interactions in young children, are cases of participatory sensemaking (De Jaegher and Di Paolo 2007). According to this concept, novel meaning is jointly achieved through a history of breakdowns and recoveries of interactive coordination. From a systemic perspective, this is a way of resolving the apparent contradiction

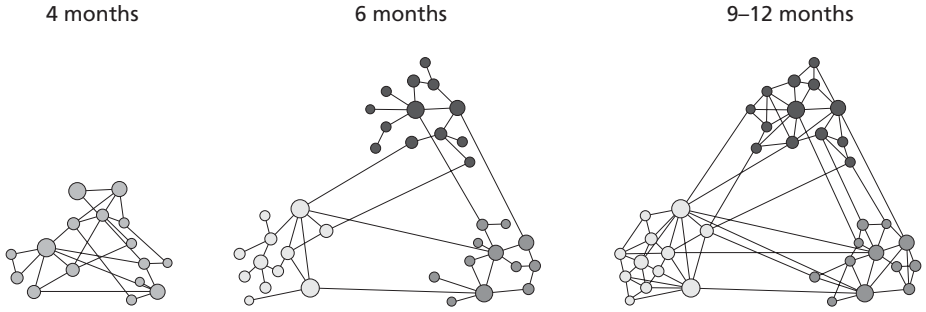

Figure 8.2 Schematic (hypothetical) representation of the scaffolding effects. These correspond to mother-infant interactions on the infant's sensorimotor network during the first year. The first panel shows the highly synergistic schemes that infants of about 4 months of age bring to an object. As the mother solicits bids for attention to herself and other objects, these schemes begin to diversify (middle panel) into new, loosely linked clustered schemes. The mother now encourages integration of these schemes during play by sustaining attention and manipulations on single objects and on relations between objects and persons. This results in the network growing increasingly consolidated as the schemes for attention, object manipulation, shared attention, social acts, etc., get progressively integrated (third panel). (See color plate.)
Copyright © 2017 Ezequiel A. Di Paolo, Thomas Buhrmann and Xabier E. Barandiaran, with permission.

between the autonomy of the cognitive, living agent and the possibility of socially constituted, incorporated skills. Social processes can intervene, but not in a determining fashion, within the self-constituted identity of the autonomous agent (otherwise she would cease to be autonomous). However, a history of co-specifications can lead to forms of individual agency that would be unreachable through the systemic transformations available only to the isolated agent.

In order to understand how these social influences transform sensorimotor agency such that objects may be perceived in the abstract, in the construct mode, we need to ask what it means for a perceptual attitude to be constituted intersubjectively. What aspects of such an attitude (which undoubtedly belongs to a particular, individual embodied agent) can be said to be essentially social?

Social interactions tend to have a life of their own and involve phenomena that are irreducible to the sum of the individual actions and intentions of the participants (De Jaegher and Di Paolo 2007; De Jaegher, Di Paolo, and Gallagher 2010). The interactive order is what brings people together (there is a pull toward sustaining interaction), but it is also a source of heteronomy for each participant. Individual acts during the interaction are subject to a double normative framework: they are, on the one hand part of individual sense-making and on the other, they are moves in the unfolding of the interactive encounter.

Cuffari and colleagues (2015) discuss this primordial tension between individual and interactive norms. Their dialectical model explores the concept of participatory sense-making, which moves from general coordination of intentional activity during interaction, through various forms of complex social agency, to "languaging." One form of self-differentiation in participatory sense-making is the emergence of complementary

social acts, acts that require interpersonal coordination to bring them to completion, such as a handshake or the act of offering/accepting an object. Each individual enacts a partial sensorimotor scheme, and to learn to perform social acts successfully implies interpersonal equilibration between these schemes. The ability to participate in such acts is an irreducibly social skill. Social interaction is not merely enabling of this skill (via mutual equilibration) but it is also constitutive because the social act can only be conceived as being bound by the normativity of the interactive order. I may perform a dexterous gesture offering you the object you have been looking for, but if I do it with bad social timing, say, when you turn to address someone else, the social act is awkward or it simply fails.

The partial schemes that jointly make a social act also tend to be reciprocal and not just mutually fitting, but also mutually addressed. At some point the child's sensorimotor repertoire is endowed with coherent partial schemes, easy to coordinate with one another (e.g., a scheme for offering an object and the scheme for receiving it), but which unlike other sensorimotor schemes, are inherently oriented toward an other (like one half of a handshake). In our sensorimotor network description, these could be represented as schemes with open arrows leading to complementary schemes in another person. But the child learns to act on both sides of the complementary relation. As the child learns to hand over a toy to the mother and then to accept it back from her, she is not too far from being able to "hand it over" to a resting place, while doing something else, and then "accepting" it again. This is different from simply putting the toy down and then picking it up. The attitude toward the toy is tinted with social normativity involving an implied other. The toy is partially objectified by the solitary combination of two complementary partial schemes enacted by a same individual. The resulting whole social—though solitary—act implies the object is being "held" and regarded by an absent viewpoint. In this way, the reciprocal sensorimotor skills involved in partial schemes invite the shifting of perspectives, which even in solitary play can be enacted, as in the above example, by the child "giving" the toy to an absent other whom she partially incarnates.

Phenomenologically, abstract object perception also points to a constitutively intersubjective connection. The abstract perceptual attitude involves perceiving the object with its properties and relations to the perceiver and its surroundings, but neutralized of any instrumental interest (as far as possible; we never reach a fully abstract attitude). Successful instrumental actions are complete in themselves. A contemplative attitude toward the object is open instead. We do not know in principle where it may lead, how long it will keep us occupied, and, if the object is unfamiliar, what we are likely to discover. The child experiences such unfinished, prolonged, sensorimotor engagements during interaction with others, in particular during play. Actions in the interactive context are hardly ever closed on themselves since they also serve in sustaining the encounter (when they fully close, they signal a pause or a break from the interaction). Playful interactions are especially open-ended; they engage the zero mode as meanings can break loose from direct instrumental concerns.

The contemplative perceptual attitude is phenomenologically afforded by the infant's experience in expecting actions not to finish with a clear goal. But this is not the only phenomenological clue. As the child regards an extraneous object, her attitude is one of curiosity, an interrogative attitude. The present object is addressed as a social other, one that is unable to complete fully the child's open partial acts. The object is interrogated with hands, eyes, mouth, and ears. And the child is open to being surprised (or disappointed or quickly bored). As she has learned in social encounters, she expects answers and solicitations to further action, but in this case, she has to facilitate them herself.

A similar interpretation is offered by Thomas Fuchs (2012) who presents an analysis of the development of first-, second-, and third-person perspectives in infancy focusing on the emergence of reflexive, meta-perspectival stances in the child that are constitutive of social and self-understanding. He highlights the developmental relevance of conflict during interpersonal interactions. Vasu Reddy (2008) suggests that there are primordial aspects of second-person interactions from which concepts of self and others develop. The infant experiences in herself what it means to be an object of attention for others and this experience serves her to understand that others may attend also to other objects and that objects can be made into social actors. This knowing oneself through another's actions and expressions fits with Merleau-Ponty's description of the difference between concrete and abstract movement. In concrete movement, the motor project is directed toward others and the world. In abstract movement, instead, it is directed toward the body itself (e.g., while trying to execute some prescribed series of postures or moves without immediate significance, the motor project aims at moving my limbs, my head, etc., as specified) (Merleau-Ponty 1945/2012, p. 114). But again, this reflexive orientation toward the body is achievable by first experiencing the body as an object toward which it is possible to hold motor projects, guided, as it were, by the attitudes of others as they direct their attention and actions to myself.

According to our enactive interpretation, as decentered attitudes develop, the child approaches objects starting from a second-person perspective; one, however, that never quite reaches the full engagement of this perspective as in a real social interaction, or does so partially (personalization of toys, animism). This perspective does not have all the elements of a proper second-person attitude as the object shows itself not to be an autonomous other. This invokes the child's use of complementary partial schemes to sustain the encounter in a quasi-interactive mode involving perspective shifting. The non-instrumental object invites and at the same time resists a second-person engagement. By eliciting in the child a second-person stance that cannot succeed, but invites creative completion in the zero mode through the adoption of multiples roles, the interrogative engagement with the object gives rise to a novel viewpoint. Out of a breakdown of the second-person attitude, a third-person perspective emerges.

8.8 **Perceptual mastery … once again**

The perspective on virtual actions and on intersubjective factors that constitute some perceptual attitudes sketched in the previous sections will require further elaboration.

Ideally, we would like to take these ideas to the operational/dynamical descriptions we have developed for other proposals in this book. These future developments will help us clarify the precise relations between the different modes of the human mind. Nevertheless, we can already re-examine some of the central aspects of the sensorimotor approach to perception based on these incomplete proposals.

We return to Noë's (2004) account according to which I experience a tomato as having an unseen side and volume, because of my "knowledge" of how sensory and motor changes would co-vary if I were to look at the tomato from a different perspective. Put in these terms, this unperformed act is just abstractly counterfactual and invites representational explanations. Gallagher (2008) rightly points out that if it were a question of imagining these non-actualized moves, I would not then perceive the hidden aspects of the object as coexisting at this moment with my view of it. I would *know* the hidden sides "as a necessary consequence of a certain law of the development of my perception" (Merleau-Ponty 1964, p. 14). But what is given in perceptual experience are not necessary truths, but concrete presences (Merleau-Ponty 1964). In other words, presence itself *cannot* be constituted by any knowledge about lawful co-variations.

Gallagher seeks the solution to the phenomenological conundrum in open intersubjectivity. According to the Husserlian account, "if the absent profiles cannot be correlated with my possible but non-actualized perceptions, then the absent profiles may be correlated with the possible perceptions that others could currently have" (Gallagher 2008, p. 172). Possible truths, however, have the same inconvenience as necessary truths: they are intellectual notions that differ from the concrete experience of presence. Those "possible perceptions" belonging to others may not be actualized either at the moment I decide to stare at the tomato. And if they were, the right synergy between my body and theirs could still be absent. They also remain abstract. In the absence of direct actualizations, these possible perceptions of others must also be imagined, remembered, or deduced. The same criticism Merleau-Ponty would make to Noë seems to apply to Gallagher's proposal. Moreover, if there is an apperceptive horizon of possible experiences subtended by the object, which includes mine and those of others, this phenomenological insight does not yet provide an explanation of how this quality of perception could come about not only in the absence of co-present alternative acts, but also without co-present others.

Gallagher is correct to invoke the intersubjective dimension so as to add degrees of freedom to the problem and aim for a closer phenomenological match. And as we have discussed, empirical evidence strongly backs this move. But he is too quick in ruling out virtual aspects of sense-making. These are not temporally disjointed from the perceptual act; in fact, we have already seen that they always accompany it. Dynamically speaking, alternative engagements with the object can be virtually co-present if we see perception as time-extended (Froese and Di Paolo 2009). However, and importantly, this occurs *in a concrete context-dependent manner*, not as conveniently invoked counterfactual acts. "The hidden side [of an object] is present in its own way. It is in my *vicinity*" (Merleau-Ponty 1964, p. 14, emphasis added). Virtualities, unlike abstract counterfactuals, are adjacent to actuality; they are in the concrete "vicinity" of the actual. They can take the form of

sensitivities to proximate situations (the risks and opportunities I confront). They also take the form of inner tensions as certain evoked sensorimotor schemes are inhibited in favor of others. These tensions are contextual in that they follow the internal relations of schemes clustered around a given activity. They may be resolved through full-fledged virtual actions, as we illustrated in Figure 8.1.

The problem is that these virtualities may or may not involve those precise sensorimotor co-variations required to reveal specific aspects of the object, such as having an unseen side. This will depend on the concrete activity. As we have said in the discussion on mastery at the end of Chapter 6, the visual presence of the tomato does not seem to be informed by arbitrary possible schemes in which it could be involved, but mostly by those schemes enacted in the context of "tomato-related" activities. Many of these involve grasping the tomato, moving it from one place to another, and so on. Graspable objects such as tomatoes and bottles involve a field of virtual actions that allow us to see them as having volumes, solidity, and hidden sides. But this is not the case with objects we engage differently. We do not see, say, keys on a computer keyboard in this way. A computer user rarely observes keys outside the context of a keyboard, and seldom accesses their hidden side. The virtualities present in normal engagements with the keyboard create a perceptual horizon that extends beyond the immediate visual and tactile sensations, but they do not suffice to constitute keys as having an underside. Keys in a keyboard are present as more than surfaces, but less than volumes, and this difference is related to the clustered schemes one typically uses to engage with them.

Of course, we *know* keys have a hidden side (and this is proof that knowledge itself cannot constitute presence). Following the line we have been pursuing, we do not need to invoke a representational story to account for this aspect should we decide to contemplate keys abstractly, outside the context of typing. Instead, we need a story, like the one we have sketched above, that explains how I am able to shift perspectives and adopt an abstract perceptual attitude. We have argued that the skill behind this act is intersubjectively constituted. I can see keys as objects not by mentally disassembling the keyboard, or by invoking a flat imaginary being who lives inside my laptop. I do it by manipulating my sensitivities to the virtualities of my situation, in other words, by enacting the skill of perspective shifting that allows me to establish a dialogical, even playful, relation to the object. As a result, I may now regard the keys no longer instrumentally, but by invoking a different sensorimotor repertoire, for instance, the one used when looking at other loose small flat objects such as buttons or coins. Then I can invoke a series of manipulatory schemes applicable to *these* other objects and influence my visual perception of keys as things I could virtually hold between my fingers and turn around. Such perspective-shifting skills come in degrees and relate to personal experience, e.g. photographers and painters appreciate complex patterns of color, texture, and light in ordinary objects more easily, in general, than other people. They can bring into play richer sensorimotor repertoires in their perspective shifts.

It would seem therefore that object presence is always activity-dependent according to the only non-representational candidate explanation we have so far, which relies on the

influence that sensorimotor schemes exert on each other. Each enactment (in particular, the enactment of schemes involved in perceiving an object) is colored by the regional influences in the cluster of schemes that make up an activity. The only way to alter how an object is perceived is to adopt a different perceptual attitude, that is, through the virtual act that allows it to be seen from the perspective of a different activity, or if possible, in the case of linguistic sensorimotor agents, to see it in a decentered manner.

The future challenge for the enactive approach will be to provide dynamical explanations of what is required for achieving these perspectival shifts so that a lived situation changes its meaning. Our proposal is that this requires *enacting* alterations to our current field of virtualities (i.e., altering that which we are sensitive to). This sophisticated skill, we propose, is constituted—not just learned—intersubjectively, and in this direction lies the future of the enactive approach to action and perception. A detached perceptual attitude is not so much an individual skill that I learn socially, but a social skill that I enact individually.