Miguel Segundo-Ortin and Glenda Satne

Chapter 4
Sharing Attention, Sharing Affordances: From Dyadic Interaction to Collective Information

Abstract: Cognitivist approaches to joint attention conceptualize it as a form of triangular interaction, between two agents and one object. When describing the interpersonal dimension of this triangle they frame it as a form of simulation, theorizing or both, involving representations of the other agent's mental states – representation of representations – and inferences.

In this paper, we advocate a different framework for understanding shared attention, the ecological psychology framework that understands attention through the notion of 'affordance'. Affordances are relational and not representational. They are direct relationships between agents and their environments. While some authors have pointed to the notion of 'social affordance' (Heft 2007, 2017; Rietveld and Kiverstein 2014; Moreira de Carvalho 2020) for understanding phenomena related to shared attention, the notion remains general and imprecise. The problem is that the notion is used indistinctively to refer to a number of different phenomena that involve social attention in very different ways. To address this issue, we offer an initial classification of different kinds of social affordances, from dyadic relations between agents, and different forms of triangular interactions, reciprocal and non-reciprocal, that provide direct and indirect information about common environments to one or both agents, all the way to collective affordances that lie at the basis of socio-cultural forms of life. We argue that this account is better placed than the standard cognitivist alternative to account for both shared attention and joint action in a non-cognitively demanding way. In addition, we show how these forms of shared activity are, in turn, fundamental for the acquisition of the socio-cultural norms that come to permeate human perception.

Introduction

The ecological theory of perception (Gibson 1966, 1979 [2015]) is famous for introducing two radical hypotheses. The first one is that perception is direct, that is, to perceive the environment we do not need to build and manipulate mental representations, but rather to detect (or 'pick up') the information that is available in the ambient.[1] The second is that perception is fundamentally perception of 'affordances'. Affordances are the opportunities for action that a specific environment offers an organism. According to this hypothesis, when we detect information about an object, for example, a glass, the first thing we perceive is the actions that we can carry out with it, such as picking it up, drinking from it, throwing it, etc. The thesis that perception is of affordances also implies that perception and action are complementary processes. This explains why ecological psychologists often speak of 'perception-action' in tandem.

In this context, 'shared attention', that is, the situation in which two agents are mutually aware of themselves attending to the same object, should be conceptualized as a form of 'social' or 'shared affordance.' Yet, ecological psychological literature on this topic has not been very precise when it comes to conceptualize the specific phenomenon that takes place when attention to affordances is shared. The key issue is to offer the nuanced concepts we need to sufficiently specify these phenomena, as to be able to distinguish them from other concepts that are similar and to which they are related in several ways.

Indeed, although ecological psychologists have accumulated a great amount of empirical evidence concerning the perception of affordances in sophisticated motor control tasks (Wagman 2019), some defenders of the theory consider that not enough attention has been paid to the study of the peculiarities of human perception-action. In particular, little attention has been given to the question of whether, and in that case how, social aspects of interaction and cultural norms that permeate human groups affect our perception of affordances. According to Heft (2018), ecological psychology rarely attends to the disposition that human beings have to guide their behavior according to social norms, and, therefore, "[i]t remains to be seen [. . .] whether this approach as articulated thus far can

---

1 The rationale for claiming that perception is not representational is the nature of perceptual information. According to ecological theory, perceptual information is specific with respect to the affordances of the environment, and it is relational, i.e., it occurs *between* the organism and the environment, not *within* the organism with respect to the environment (for a detailed argument about why the notion of representation is incompatible with ecological theory, cf. Segundo-Ortin et al. 2019).

adequately capture the socio-cultural dimensions of human action and experience" (124; cf. also Brancazio and Segundo-Ortin 2020; Segundo-Ortin 2020).

This paper[2] aims to explore perception-action when this is a matter of an agent relating to another agent. One central case of this is when affordances are the product of socio-cultural norms. We understand socio-cultural norms as patterns of thinking and action, both broadly understood, that have their origin in a practice that prevails in the context of a community of agents (Bicchieri 2006; Satne 2015). These norms emerge from the intersubjective evaluation by the community members of each other's actions, and "express the group's expectations for how anyone who would be one of 'us' should act, on pain of admonishment, punishment, or ostracism" (Tomasello 2019, 254). Our analysis starts by looking at how socio-cultural norms permeate human perception. It then moves forward the ecological psychology account of these phenomena by expanding on the concept of 'social affordance', i.e., what it means to say that an affordance is shared by two agents and in what different ways an affordance can be said to be shared. This analysis is then put at the service to illuminate the question of how socio-cultural norms come to permeate human perception.

In the first section, we consider the way different theorists of ecological psychology have approached theorizing on social affordances, that is, focusing on the relationship between socio-cultural norms and perception-action. In the second section, we address Tomasello's (2014, 2019) account of social learning and the acquisition of these norms by children. The analysis of Tomasello's proposal is relevant, since it is often cited by authors who have tried to address the relationship between the socio-cultural context and the perception of affordances (cf. Heft 2007, 2017). Nonetheless, we claim that Tomasello's cognitivist proposal is incompatible with adopting an ecological perspective of perception and intersubjective action, and that it is problematic in its own terms when one tries to give an account of basic forms of shared attention. Finally, the last two sections outline an alternative to Tomasello's account and present the outline of an account of social interaction, social perception and the learning of socio-cultural norms, and their mutual relations. While this proposal is presented in broad strokes, we claim that it sits well with the main tenets of ecological psychology and thus, it deserves further exploration.

---

2 A previous version of these ideas was written in Spanish for a chapter on social affordances, "Affordances y normas socio-culturales", to be included in the forthcoming collected volume *Affordances y ciencia cognitiva. Introducción, teoría y aplicaciones*, edited by Heras-Escribano, M.; Lobo, L.; Vega, J.

1 Affordances, Social Interaction and Socio-Cultural Norms

Most of our actions are mediated by socio-cultural norms. For instance, when we go to a work meeting, we look for a free chair to sit on, instead of sitting on the floor or on a table. In addition, when we sit down, we adopt a specific position, unlike the one we would adopt at a dinner with friends. Likewise, when we enter an elevator, we try to maintain an 'adequate' distance from others, and we can appreciate when other people do not do it.

As Haugeland explains:

> [W]hen community members behave normally, how they behave is in general directly accountable in terms of what's normal in their community; their dispositions have been inculcated and shaped according to those norms, and their behavior continues to be monitored for compliance   (Haugeland 1990, 440)

This fact has not gone unnoticed by some theorists of ecological psychology. For example, Heft (2001, 2007, 2017, 2018) claims that a fundamental aspect of the human condition is that our perception-action is shaped by the socio-cultural context in which we live. Thus, our relationship with the affordances of the environment does not depend only on the detection of information that is present in our natural environment but is affected by the social norms of the community of which we are part (Rietveld and Kiverstein 2014). Likewise, Reed explains that while learning to perceive the affordances of the environment requires discovering the relationship between our action capabilities and the properties of the environment, "[w]hen one learns about norms, one is learning about properties of one's own action (and their objects) with respect to the awareness and activities of others" (Reed 1993, 52).

Following Reed, Heras-Escribano writes:

> [T]he taking of affordances can be affected by certain pressures exerted by social norms. This happens constantly in our everyday situations: We do not eat with our hands for a normative reason, even when we sometimes can grasp the food more firmly with our hands than with a fork and a knife; also, someone in a hurry gives preference to an elder instead of blindly taking the affordance of passthrough-ability at the gates of a subway train. Our social norms and conventions share their space with our individual perception of affordances, and sometimes our norms exert some pressure for not taking certain affordances given some social conventions.   (Heras-Escribano 2019, 175)

Furthermore, Gibson (1950) distinguished between 'expedient' and 'proper' action. While the former refers to those actions that are useful to achieve a specific goal, the latter refer to those that are appropriate in the context of a community (Heft 2018, 126). Crucially, in most cases these two criteria do not go together. On

the contrary, according to Gibson we frequently act according to what is considered appropriate or adequate in our community or social group instead of doing what is most efficient considering our current goal (Gibson 1950, 153). Furthermore, it is often the case that social norms do not affect all members of the group in a homogeneous way. As Reed (1993, 52) suggests, human communities often have specialized roles related to gender, age, or socioeconomic status, and these roles carry restrictions on what affordances can be used, by whom, and under what circumstances.

But the existence of social norms not only influences how we relate to the affordances existing in the environment. Some affordances are also a product of these norms. To illustrate this, Gibson appeals to mailboxes (Gibson 1975 [2015], 130). According to Gibson, for a metal box located in the middle of the street or on the facade of a building to offer the possibility of sending and receiving letters, there must be a community with a postal system. Costall (1995, 2012) refers to the affordances that depend on the existence of social norms as "canonical affordances". According to Costall, these affordances only exist in relation to a shared socio-normative context, and can only be perceived by those individuals who are aware of the norms that support them.

Social norms thus play a fundamental role both in creating and in shaping our interaction with some affordances. These social norms are seldom verbalized, but they are manifested as embodied habits of perception-action that predispose us to perceive and take advantages of certain affordances instead of others and in particular situations (Heras-Escribano 2019; Segundo-Ortin 2020; Menary 2020; Segundo-Ortin & Heras-Escribano 2021). It is undeniable that human beings are in contact with socio-cultural norms from the moment they are born, but how do we learn to behave according to them? Most authors agree that it is through social interaction that we learn to coordinate our perception-action with respect to these norms (cf. Satne 2015; Krueger 2011, 2013; Reddy 2015; Tomasello 2019).

Theorists of ecological psychology have long noted the importance of social interaction to understand perception. As Gibson explains: "it is a mistake to construct a behavior theory without reference to social interaction, and then to attach it only at the end" (Gibson 1950, 155). As Heft (2007) expounds, adults often guide children's attention to objects and show them, either implicitly, through demonstrations, or in the context of cooperative actions, how to interact with those objects appropriately. According to Heft, "[s]uch intersubjective acts of 'joint attention' [. . .] contribute to the developing patterns of selection in perception-action, a process of *guided attunement*, which forms a crucial part of the child's history as an agent" (Heft 93, emphasis original).

However, these observations by Gibson and Heft have not been generally recognized by ecological psychologists. Instead, as Adolph and Hoch (2019) argue,

most ecological psychologists tend to adopt an individualistic approach when studying perceptual learning, which leads them to ignore the role that social interaction plays in this process. In the next section, we focus on analyzing the proposal of Tomasello (2014, 2019) who emphasizes the role of social interaction and joint attention in learning and the acquisition of socio-cultural norms. The figure of Tomasello is important in this regard, for it is often used as a reference by those theorists who challenge the individualistic approach to perceptual learning and aim to understand how social norms and social interaction influence human perception-action (cf. Heft 2007, 2017; Costall 2012; Rietveld and Kiverstein 2014).

2 Learning Socio-Cultural Norms

For Tomasello (2014, 2019), what distinguishes human cognition from that of other primates is the ability to participate in collaborative actions that involve common goals. According to Tomasello et al., "[t]he result of participating in these activities is species-unique forms of cultural cognition and evolution, enabling everything from the creation and use of linguistic symbols to the construction of social norms and individual beliefs to the establishment of social institutions" (2005, 675).

Tomasello explains the learning of social norms by postulating two different cognitive capacities: joint intentionality and collective intentionality. 'Joint intentionality' involves joint attention but goes beyond it. While joint attention implies the mutual awareness that both agents are attending to the same object or affordance, joint intentionality involves on the part of the agents the ability to pursue a common goal, and thus the ability to manipulate the object that is jointly attended to, with respect to the common end that both agents share. Hence, joint intentionality refers to the ability to collaborate with others in short-term face-to-face relationships. These are interactions in which the individuals involved share a common goal and jointly attend to situations or objects in the immediate environment in order to pursue said goal. Importantly, although joint attention is a common phenomenon in many primates, Tomasello thinks that joint intentionality is an exclusively human cognitive capacity (Tomasello 2019, 82).

Joint intentionality, suggests Tomasello, appears between 14 and 18 months of age, when children acquire the ability to form a "joint agent" with others (Tomasello 2014, 39; 2019, 87). It is noteworthy that in this type of collaboration, children are already exposed to social norms (Reddy 2015; Krueger 2013). However, says Tomasello (2019, 250; cf. also Hardecker and Tomasello 2017), children understand these norms as impositions or requirements from the other person (usually an

adult), and not as expectations that go beyond the concrete interaction and that apply to all members of the community.

The other great cognitive revolution occurs around the age of 3. At this point, children develop what Tomasello calls 'collective intentionality'. This type of intentionality goes beyond the immediate nature of joint intentional engagements and allows children to pursue long-term collaborations with others. It is important to point out that, while in the type of collaborations typical of joint intentionality children are capable of adopting the perspective of another agent, from the age of 3 they start "'collectivizing' [. . .] perspectives and positing a kind of invariant objectivity that grounds them all" (Tomasello 2019, 77). This ability to adopt an objective perspective is crucial for learning (and expecting others' conformity to) social norms:

> [Collective intentionality] begins around three years of age and transforms joint commitments into collective commitments, second-personal protest into the enforcing of social norms, and a sense of fairness toward individuals into a sense of justice to all in the group.   (Tomasello 2019, 251)

That children over 3 years old are capable of correcting the behavior of others for reasons that go beyond their own interest suggests that they are capable of adopting this objective perspective. For example, according to Vaish et al. (2011), children at this age show a tendency to protest when someone shows signs of wanting to break someone else's toy. As Tomasello explains, since the child is not affected by the action of the other directly, her reaction does not constitute a second-person protest, that is, a reaction in which the child seeks retribution for damage or injustice that another has inflicted upon her. On the contrary, "[w]hat she is protesting is a lack of conformity to the group-minded social norm for how one should treat others" (Tomasello 2019, 256).

As we mentioned before, Tomasello is often cited by those authors who, from ecological psychology, aspire to understand how our perception-action is affected by socio-cultural norms. Nevertheless, our thesis is that Tomasello's proposal is in direct conflict with an ecological approach to perception and social action and unsuited in its own terms to give an account of the basic forms of shared attention that lay at the basis of the acquisition of socio-cultural norms.

To see this, we must take a closer look at the way Tomasello understands joint intentionality, the previous and necessary step for the development of collective intentionality. Tomasello (2019, 7; 2014, 38) makes it clear that he understands joint intentionality in terms of the theoretical framework proposed by Bratman (1992, 2014). Accordingly, for there to be joint intentionality, the following three conditions must be satisfied:

If you and I are agents, and *J* is a goal, then:
(1) I must have the goal of doing *J* together with you,
(2) You must have the goal of doing *J* together with me, and
(3) We must have "mutual knowledge, or common ground, that we both know each other's goals" (Tomasello 2014, 38)

According to Tomasello, to account for this sort of social interaction, we must explain how it is possible for an agent to know that another agent knows that they have the goal of carrying out *J* with them. This, he argues, is only possible if the agents have the ability to recursively read each other's minds (Tomasello 2019, 85). For this, I must be able to represent the mental states of the other (in this case her intentions and beliefs), simulate the abductive inferences that she is making about my intention to carry out *J* with her, and infer that she knows the same as I do about the situation in which we are to *J* (Tomasello 2014, 94). All of this involves according to Tomasello, the ability to manipulate recursive inferences indefinitely iterated.[3]

In summary, to explain joint intentionality we have to assume: (i) that the agents involved possess concepts such as 'intention,' 'belief,' 'desire', etc.; (ii) that they are capable of forming representations about the mental states of other agents; (iii) that they are capable of carrying out complex cognitive operations such as simulating the inferences that the other is making about their intention to carry out J with her; and (iv) that they have the ability to know that the other agent knows what they know, by indefinitely iterating recursive inferences to that effect.[4]

Nonetheless, if we take into account that, according to Tomasello, the ability to form joint intentions appears between 14 and 18 months, it seems excessive to assume that children of that age are already capable of carrying out cognitive operations of such complexity. In fact, several authors have argued that attributing these capabilities to young children is implausible (Tollefsen 2005; Michael et al. 2014; Pacherie 2013). These authors propose, alternatively, that the ability to engage in joint actions with others could appear in infancy *before* the development of such complex recursive cognitive abilities (Satne

---

3 Tomasello explains these capacities by combining Theory-Theory and Simulation Theory.

4 Tomasello suggests that some situations offer enough information for the two agents to understand that they have a common goal without the need for recursive inferences, standing in 'common ground'. Yet, according to him, situations in which agents do in fact make those inferences, for example when there is a potential misunderstanding, demonstrate that agents do possess those underlying recursive inferential capacities, and that those can be legitimately presupposed as explanations of jointness (Tomasello 2014, 38).

2016; Satne and Salice 2020). Continuing with this line of thought, some theorists argue that it is precisely the interactions with other agents, especially those interactions that involve language and narratives, that provide children with the cognitive tools that allow them to grasp concepts such as 'intention' and 'belief'. This also allows them to represent other agents the mental states, of increasing complexity, through recursive inferences (Hutto 2008; Hutto and Satne 2015). This means, against Tomasello, that we should be able to explain joint intentionality in a simpler way, without resorting to handling representations of concepts for mental states and recursive inferences from the set off.

Moreover, Tomasello's proposal is in direct conflict with ecological psychology. Although Tomasello (2000; Tomasello et al. 1999) uses the term "affordance" on numerous occasions, and even refers to Eleanor J. Gibson's work on perceptual learning, he assumes that we cannot understand other agents' mental states through direct perception. In contrast, James Gibson (1979 [2015], 127) argues that we can perceive the affordances of other agents in the same way we perceive the affordances of objects: detecting directly, without mediation of inferences or 'theories', sensory information in the environment.[5] Therefore, ecological psychology argues that we do not need to postulate the existence of meta-representations and abductive inferences to explain our ability to understand, at least in basic cases, what others are doing and to cooperate with them to carry out common goals.

From this we can conclude two things. First, that Tomasello is not the best ally for authors, like Heft and Costall, who, from an ecological psychology perspective, try to explain how children learn social norms through social interaction. Tomasello's account of social cognition builds on the notions of meta-representation and recursive inference, something that does not sit well with ecological psychologists' claims about the direct nature of perception (including social perception). Second, ecological psychology's account of perception in terms of affordances, might be in the position to offer novel resources for elucidating how children interact with others and learn collective norms without resorting to an explanation that is excessively cognitively demanding, like the one provided by Tomasello. With this in mind, in the next section we distinguish different types of 'social' affordances. These will be shown to be a crucial set of concepts to understand the phenomena of shared attention in the context of ecological psychology. In the last section, the distinction of 'social affordances' in different kinds will prove to be useful to address the question of how children learn of socio-cultural norms. In contrast to what Tomasello's account assumes, such learning turns out not to be such a cognitively demanding endeavour.

---

5 As we will discuss below this is one case of what we can call 'social affordances'.

3 From Social Interaction to Social Affordances, and Back Again

In a trivial sense, we could say all human affordances have a social character (Costall 1995). This is because, as we have said before, our perception-action of the environment is mediated by the social norms of the community we partake in (Rietveld and Kiverstein 2014). Furthermore, it is almost impossible to find an environment that has not already been transformed by the action of human beings, and that does not bear the "stamp of the social" (Heft 2007, 95). However, in order to advance a more nuanced characterization of the role that social affordances play in perception-action, in this section we will propose a more restricted characterization of the concept of 'social affordance,' as well as distinguish between different types of social affordances.

We can characterize social affordances as those opportunities for action that depend on the presence in my environment of one or more agents with whom I can interact. Following Marsh et al. (2006, 2009a, 2009b), we defend that a study of social affordances requires taking as the minimum unit of reference the O-O-E [Organism-Organism-Environment] system, instead of the classic O-E [Organism-Environment] system of ecological theory. Our aim is to show how an ecological approach to social interaction based on the concept of social affordance can account for joint intentionality without the need to postulate meta-representations and recursive inferences. According to this, our hypothesis is that joint intentionality can be understood as the coordinated exploitation (by two or more agents) of the affordances of the environment, and that this is possible through the direct perception of social affordances.

To begin with, we must note that the environment we inhabit is often populated by other agents, and these agents offer affordances we can perceive:

> The richest and most elaborate affordances of the environment are provided by other animals and, for us, other people. [. . .] Behavior affords behavior, and the whole subject matter of psychology and of the social sciences can be thought of as an elaboration of this basic fact. Sexual behavior, nurturing behavior, fighting behavior, cooperative behavior, economic behavior, political behavior – all depend on the perceiving of what another person or other persons afford, or sometimes on the misperceiving of it.
>
> (Gibson 1979 [2015], 126–127)

But in what sense does the existence of other agents imply the existence of 'richer' affordances? This could be due to several factors (Reed 1993; Gibson 1979 [2015], 127). First, other agents are not passive. Agents *act*, transforming the environment and generating new affordances for others. For example, it may happen that while we are walking along a busy street, a distracted pedestrian changes

her trajectory, approaching us in a straight line, forcing us to avoid her. It can also happen that this same person starts running towards us, aggressively waving her arms, forcing us to flee from her. Second, when we interact with other agents, they can either resist or collaborate with us. Two people can transport an object together, coordinating their movements and their attention, or they can compete to see who carries the object. Both cases imply totally different interactions, and with them, different affordances that can be perceived and used by each of the agents. Furthermore, as Baron (2007) suggests, it often happens that when we collaborate with others we tend to adopt different roles, and these roles determine what affordances are immediately relevant for us.

In what follows, we distinguish four types of social affordances, according to the type of interaction involved and the factors that need to be taken into account to describe each one. These distinctions allow to enrich the concept of social affordance and make it useful to describe various types of perception-action within the O-O-E system.

As Gibson (1979 [2015], 127) suggests, just as there is information in the environment about the affordances of objects, there is also information about the affordances of other organisms. Following Reed (1993), we propose that the most basic case of social affordance is that in which an agent (O_1) is able to perceive the presence of another organism (O_2) by detecting some traces that it leaves in the environment (Figure 4.1). These traces can constitute perceptual information about other organisms, and their detection is essential for the control of action. This type of perception-action is common in the animal kingdom. Some predators are able to follow the trail of their prey by perceiving the smell they leave, for example, and dogs and wolves can determine the proximity of the prey based on the intensity of the smell.
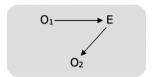

Figure 4.1: The first agent (O_1) perceives the presence of another agent (O_2) detecting some type of trace that it left in the environment (E).

Another kind of case is one in which an agent perceives an affordance of the environment indirectly, that is, through the *action* of another agent (Figure 4.2). For example, it is well known that the primates of the species *Chlorocebus pygerythrus* emit a characteristic sound to alert their conspecifics about the presence of predators. These 'alert calls' are also different depending on the type of predator they identify. Perception of these sounds generate differentiated responses in the

other members of the group (Seyfarth et al. 1980). It is important to note that, although this type of situation depends on cooperative forms of interaction between several organisms, we cannot speak of 'joint' action yet, since organisms performing the calls do it even when they are not aware of the presence of others receiving the calls.
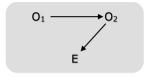
Figure 4.2: O_1 perceives an affordance of the environment (E) by means of O_2.

In the case of humans, direct perception of other organisms' actions is pervasive. Johansson (1973) filmed a series of people performing different physical activities (running, walking, dancing, lifting objects, etc.) in the dark while having different light points attached to their joints. Afterwards, he showed the videos to a series of participants and found that a high percentage of them were able to identify not only the type of activities they were doing, but also how much effort it was taking them to do so (cf. also Runeson and Frykholm 1983). Subsequent experiments show that people can perceive whether the other person intends to carry out this or that action – that is, to exploit this or that affordance – observing their movements as well as whether the movements are performed with the awareness that others are observing them (Runeson 1985; Hodges and Baron 2007; Mark 2007).

In the same vein, Kiverstein (2015) points out that the ability to detect which affordances are relevant and significant for an agent (Figure 4.3) allows humans to become aware of the other agent's mental states (e.g., if she is angry, if she has an aggressive attitude, or if she intends to cooperate with us, etc.) directly (not inferentially). In the words of Gallagher and Hutto,

> in most intersubjective situations, that is, in situations of social interaction, we have a direct perceptual understanding of another person's intentions because their intentions are explicitly expressed in their embodied actions and their expressive behaviors. This understanding does not require us to postulate or infer a belief or a desire hidden away in the other person's mind.   (2008, 20; cf. also Krueger 2011; for an overview cf. Satne 2020)
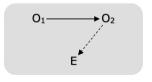
Figure 4.3: O_1 perceives O_2 acting in the environment (E). In this case, the action of O_2 provides information about the mental states of O_2. The dashed arrow represents the action of O_2 on E.

According to Marsh et al. (2006, 2009a, 2009b), "direct social perception" – the name by which they refer to our ability to perceive the intention of the other through their action – is essential for interpersonal coordination. Once I am aware of the intention of the other, I can decide whether or not to collaborate with her. When this collaboration occurs, Marsh et al. suggest, a new unit of perception-action is created – a "plural subject of action" (Richardson et al. 2007) or a "joint agent" in Tomasello's words –, meaning that my perception-action is coordinated with that of the other to achieve a common goal:

> Just as perception and action are mutually and causally coupled to behavioural aims at the individual level – by the detection of information to constrain action and by the control of action to order perception – the perception and action capabilities of the social unit are mutually constrained, ordered, and dynamically coupled [. . .] Each individual's perception is coupled to his or her partner's action as it is to his or her own, and each individual's action alters their partner's perception just as it alters his or her own [. . .] the perceiving and acting of those individuals within the social unit are causally entailed to form a distinct but irreducible system motivated by a mutually perceived goal.
>
> (Marsh et al. 2006, 20)

To illustrate this idea, Marsh et al. (2006) asks us to imagine two people carrying an object, for example a table. To carry out this action, both individuals have to coordinate their perception-action with respect to both the affordances of the environment and the action of the other (Figure 4.4). The key for this to be possible, they explain, lies in the detection of information that is generated at the level of the O-O-E system (information about social affordances), and not in the knowledge and simulation of the inferences of the other agent (Marsh et al 2006, 22; Hodges and Baron 2007). Through this interpersonal coordination, agents can carry out a common task, exploiting the affordances of the environment in a coordinated way (Marsh et al. 2009a, 2009b; Baron 2007). Something similar, says Reed (1993, 58), occurs when, through gestures, we call the attention of other agents to a specific aspect of the environment. In both cases, the perception of an affordance of the environment is mediated by the perception of the action of another agent and the agents must coordinate their responses for the interaction to be successful.
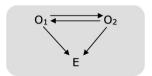
Figure 4.4: To carry out a joint action, O_1 and O_2 have to coordinate their perception-action with the environment (E) and the action of the other simultaneously.

In sum, we hold that the ecological perspective offers resources to understand social interaction without the need to postulate complex cognitive processes as Tomasello does (2014, 2019). According to the analysis offered in this section, the notion of social affordance (in the multiple ways in which these are presented – cf. Figures 4.1–4.4) allows us to account for how it is possible for an agent to perceive the possibility of interacting with others without the need to simulate and infer their mental states. It follows that the ecological theory of perception-action can help to explain the type of cognitive activities characteristic of joint intentionality without assuming that it depends on abilities such as those of representing mental states and making inferences on the basis of these. Once we have this firm foundation to start from, we can see how socio-cultural norms are learned. This is addressed in the next and final section.

4 An Ecological Approach to the Learning of Socio-Cultural Norms

In the previous section, we have shown that the ecological theory of perception-action offers resources for explaining joint intentionality without resorting to meta-representations and inferences. It now remains to be seen how to articulate a vision of the learning of socio-cultural norms compatible with this account. In this section, making use of studies from developmental psychology, we offer some suggestions about the different cognitive abilities that could contribute to the learning of socio-cultural norms. Although the capacities we call upon in what follows are not standardly studied in the ecological theory, we believe that they are perfectly compatible with its fundamental tenets, and, therefore, that they can contribute to extend the ecological approach of perceptual learning to account for the learning of socio-cultural norms. It should be noted that these suggestions do not constitute a complete account of such learning, but rather a platform along which such a theory could be developed.

As we mentioned before, children are in contact with normative social practices from birth. For example, caregivers often carry out actions (gestures, facial expressions, sounds, etc.) with the aim of regulating children's perception-action patterns of response. These actions, Krueger argues, "encode the norms, values, and patterned practices distinctive of their specific socio-cultural milieu [. . .]. These physical interventions are thus arguably the earliest examples of social practices that scaffold the infant's cognitive development and shape the development of their cultural education" (Krueger 2013, 40).

Following the reasoning from the previous section, our hypothesis is that the ability to learn from others in the context of these normative practices arises in development before the ability to form mental representations and make inferences about the mental states of other agents (Satne and Salice 2020; Hutto and Satne 2015; Satne 2016; Gallagher and Hutto 2008). We hold that the perception of social affordances is key to account for these early forms of cultural learning.

Some authors have pointed out that the ability of human beings to dynamically coordinate our movements with those of others constitutes a fundamental aspect for learning social norms (Pacherie 2013; Knoblich et al. 2011). Empirical evidence suggests that this capacity could be innate, present in the so-called "primary intersubjectivity" (Trevarthen 1979), that is, the interactive capacities that a child has from birth on, and almost exclusively, until 6 months of age. For example, babies between 2 and 4 months can identify when an adult intends to pick them up, and adapt their body posture to the way a specific caregiver picks them up even before contact (Reddy et al. 2013; Reddy 2019). If we take into account that the way in which caregivers take children is already governed by socio-cultural norms (Krueger 2013; Reddy 2015), the fact that children are able to coordinate their perception-action to the behavioral habits of their caregivers can be considered a first form of conformity, albeit very basic, to social norms.

As we mentioned in the previous section, the studies by Johansson (1973) and Runeson and Frykholm (1983) suggest that we can directly perceive the intention of another agent to carry out an action by detecting specific patterns in their body movements. This suggests that social interaction could be based on the perception of affordances of other agents (social affordances in our terminology), without the need to carry out abductive inferences or use meta-representations of the mental states of other agents.

Another capacity also present in primary intersubjectivity is to perceive the emotional responses that others have to our actions and adapt to them. This capacity is observed as early as 2 months of age, when children co-ordinately react to the facial expressions of their caregivers (Trevarthen 1979; Reddy 2019; Gallagher 2013). As Adolph and Hoch explain, "infants and caregivers are acutely sensitive to each other's facial gestures and vocalizations and use this social information to update their own actions in real time" (Adolph and Hoch 2019, 26.11). By being able to detect the emotional responses that their actions produce in adults, children begin to acquire a basic knowledge of what types of actions are acceptable in the specific context of that interaction (Gallagher 2013; Kiverstein 2015; Satne 2014). Thus, both the action of others, as well as their gestures, offer sensory information about affordances that the child can perceive and use to coordinate their action (Figures 4.2a and 4.3a)
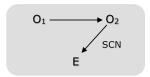
Figure 4.2a: O_1 perceives an affordance of the environment (E) by means of O_2. The action of O_2 is already mediated by socio-cultural norms (SCN), meaning that the perception of E by O_1 is influenced by these SCN.
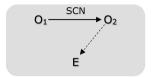

Figure 4.3a: O_1 perceives O_2 acting in the environment (E). The action of O_2 provides information about the mental states of O_2. By perceiving the mental states of O_2, O_1 begins to learn what kinds of actions are acceptable and which are not, which implies a learning of social norms (SCN).

Joint attention also plays a fundamental role in learning social norms. This involves the mutual awareness by the two agents of the presence and features of an object in the environment they are both attending to. According to developmental psychologists, this ability is consolidated between 6 and 9 months of age, when children are able to follow the gaze of their caregivers towards specific aspects of the environment and inquire back into the adults' gaze while exploring their meaning. This ability becomes more acute as the children interact with the contextual use of gestures and vocalizations by adults. Through these interactions, adults educate the child's attention, allowing them to identify those affordances that are relevant for a specific purpose (Krueger 2013; Tomasello 1999, 2000). As the child grows, she no longer needs an adult to guide his attention in order to perceive and respond to these affordances and begins to take the initiative by indicating her objects of interest to others, for example by using his fingers to point at something, or her gaze to guide the gaze of others. Later, children begin to use words to draw their caregivers' attention to something they want. For example, it is common for Spanish-speaking young children to use words like 'water', to indicate food and drink interchangeably. These types of interactions are already part of the so-called "secondary intersubjectivity" (Trevarthen 1979).

Habitualization to these patterns of perception and action is key to children's development in this stage. The child is already adapting and learning social norms by means of interacting with her caregivers and the cultural material

environment that surrounds her. These norms, patterns of thought and action, are manifested as embodied habits of perception-action that predispose the child to perceive and take advantage of certain affordances instead of others and in particular situations (Heras-Escribano 2019; Segundo-Ortin 2020).

In secondary intersubjectivity,[6] children learn from adults how to place themselves in pragmatically defined contexts, that is, specific social practices, that embody aims, goals, action-styles, etc. At this point, children begin to incorporate more explicit social norms to the ways they behave – for example, they learn how to use the spoon and the plate, or chopsticks, to feed themselves (cf. Figure 4.4a). The ability to synchronize attention with other agents continues to be key in this stage of their development, and imitation in this period becomes pervasive (Rochat 2012).

Several studies conducted by Tomasello and his team (Tomasello 1999, 2019) show that while chimpanzees are capable of emulating the behavior of others, for example, they are capable of replicating the behavior of others for a specific purpose when they have seen others do it, children tend to imitate the behavior of others, that is, they tend to replicate even arbitrary features of observed behavior. Through imitation, children incorporate specific perception-action patterns that are already fully subject to social norms (Rochat 2015), and begin to expect others to conform to them at around 3 years of age (Rackoczy and Tomasello 2012). In addition, it is important to note that, although imitation does not necessarily imply joint action, it does involve joint attention: the child needs to pay attention both to the affordances of the object and to the way in which the adult interacts with them and, in order to regulate her action for acceptance and adequacy, the child uses adult approval and disapproval as a guide (Satne 2014), which implies mutual awareness of responding to the same affordances, and the aim to do it in the same way (cf. Figure 4.4a below).

Furthermore, studies by Gergely and Csibra (2009) show that only when learning is accompanied by express instructions, this is when observation is accompanied by interactions in which the adult guides the child using language, children generalize the learned perception-action patterns to other contexts. Gergely and Csibra call this form of learning "natural pedagogy", and suggest that it is both innate and fundamental to learning social norms.[7]

---

6 There is no consensus regarding the age at which secondary intersubjectivity starts, but it is generally accepted that its appearance coincides with the emergence of joint attention. While Reddy (2019) places this at 6 months of age, for Tomasello (2019) joint attention does not appear until approximately 9 months.

7 Gergely and Csibra (2009) explain this ability in representational terms. A direction that we resist, as explained above. We suggest that the ability to follow linguistic instructions that

The following scheme (Figure 4.4a) illustrates how socio-cultural norms transform the relation between the observers and with their (common) environment:
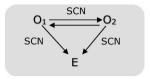

Figure 4.4a: O_1 and O_2 coordinate their perception-action with the environment (E) and the action of the other simultaneously. In this case, the perception-action of both agents is subject to socio-cultural norms (SCN) that determine what affordances are relevant in this specific context.

In sum, although we agree with Tomasello that social interaction, joint attention and joint action are key to learning social norms, our hypothesis is that these can be explained through the direct perception of social affordances and the mutual awareness of common affordances, without the mediation of meta-representations and inferences about the mental states of other agents.[8] We agree with Tomasello (2019, 88) also that linguistic interaction is basic to develop the collective perspective characteristic of social norms. However, we argue that the ability to form representations, be them of the world or of other agents, is not part of the basic cognitive repertoire that allow children to acquire such cultural perspectives. Contrary to the position defended by Tomasello, we hold that it is only when children master the socio-normative practices associated with the use of language that they have the resources necessary for forming objective representations about the world and other individuals (Hutto and Myin 2013, 2017; Hutto and Satne 2015).

---

these authors discuss can be described as a capacity to attune to perception-action patterns in interactive situations, and not as a representational inferential capacity that targets other people's mental states.

8 Note that this idea is in line with recent claims by Heft that joint and collective actions require common awareness of mutual goals and affordances, but that such common awareness "does not require mind-reading, of any sort, but instead what is needed to bootstrap joint and collective processes is access to a common ground for perception-action. [. . .] knowers have a common ground of information" (Heft 2019, 202). The aim of our proposal is precisely to give a sufficiently rich description of what the building of such common ground might consist in in the absence of mind-reading capacities.

5 Concluding Remarks

In this chapter, we have defended not only that joint action is possible without mental representations, but that learning the socio-cultural norms that permeate our perception-action as a whole, does not depend on the existence of such representations. Rather, this socio-cultural learning depends on our habitualization to embodied socio-cultural norms that we learn from other community members from early childhood.

The ecological perspective, we have argued, offers through the notion of social affordance, in its different types, fundamental tools to make sense of this possibility. By repeatedly engaging in social interactions that involve the mutual coordination and perception of social affordances, children learn social norms, consolidating perception-action habits that enable them to behave according to what is considered normal or acceptable within their communities.

References

Adolph, K. E.; Hoch, J. E. (2019): "Motor development: Embodied, Embedded, Encultured, and Enabling". *Annual Review of Psychology* 70:1, 141–164.

Baron, R. M. (2007). Situating Coordination and Cooperation Between Ecological and Social Psychology. *Ecological Psychology*, 19(2): 179–199.

Bicchieri, C. (2006): *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Brancazio, N.; Segundo-Ortin, M. (2020): "Distal Engagement: Intentions in Perception". *Consciousness and Cognition* 79, 102897. https://doi.org/10.1016/j.concog.2020.102897

Bratman, M. (1992): "Shared Co-Operative Activity". *Philosophical Review* 101:2, 327–341.

Bratman, M. (2014): *Shared Agency: A Planning Theory of Acting Together*. Oxford: Oxford University Press.

Brownell, C. A. (2011): "Early Developments in Joint Action". *Review of Philosophy and Psychology* 2, 193–211.

Costall, A. (2012): "Canonical Affordances in Context". *Avant: Trends in Interdisciplinary Studies* 3:2, 85–93.

Costall, A. (1995): "Socializing Affordances". *Theory & Psychology* 5:4, 467–481.

Csibra, G.; Gergely, G. (2009): "Natural Pedagogy". *Trends in Cognitive Sciences* 13:4, 148–153.

Gallagher, S.; Hutto, D. (2008): "Understanding Others through Primary Interaction and Narrative Practice". In: J. Zlatev; T. Racine; C. Sinha; E. Itkonen (eds.): *The Shared Mind: Perspectives on Intersubjectivity*. New York, NY: John Benjamins, 17–38.

Gallagher, S. (2013). The Socially Extended Mind. *Cognitive Systems Research*, 25: 4–12. https://doi.org/10.1016/j.cogsys.2013.03.008

Gibson, J. J. (1950): "The Implications of Learning Theory for Social Psychology". In: J. G. Miller (ed.): *Experiments in Social Process: A Symposium on Social Psychology*. New York, NY: McGraw-Hill, 149–167.

Gibson, J. J. (1966): *The Senses Considered as Perceptual Systems*. Westport, Conn.: Greenwood Press.

Gibson, J. J. (1979[2015]): *The Ecological Approach to Visual Perception*. New York, NY: Psychology Press.

Hardecker, S.; Tomasello, M. (2017): "From Imitation to Implementation: How Two- and Three-Year-Old Children Learn to Enforce Social Norms". *British Journal of Developmental Psychology* 35:2, 237–248.

Haugeland, J. (1990): "The Intentionality All-Stars". *Philosophical Perspectives* 4, 383–427.

Heft, H. (2001): *Ecological Psychology in Context: James Gibson, Roger Barker, and the Legacy of William James's Radical Empiricism*. New York, NY: Psychology Press.

Heft, H. (2007): "The Social Constitution of Perceiver-Environment Reciprocity". *Ecological Psychology* 19:2, 85–105.

Heft, H. (2017): "Perceptual Information of 'an Entirely Different Order': The 'Cultural Environment'". The Senses Considered as Perceptual Systems". *Ecological Psychology* 29:2, 122–45.

Heft, H. (2018): "Places: Widening the Scope of an Ecological Approach to Perception–Action with an Emphasis on Child Development". *Ecological Psychology* 30:1, 99–123.

Heft, H. (2019): "Revisiting 'The Discovery of the Occluding Edge and its Implications for Perception' 40 years on". In: J. B. Wagman; J. C. Blau (eds.): *Perception as Information Detection. Reflections on Gibson's Ecological Approach to Visual Perception*. New York, NY: Routledge, 188–204.

Heras-Escribano, M. (2019): *The Philosophy of Affordances*. Cham: Palgrave Macmillan.

Hodges, B. H.; Baron, R. M. (2007): "On Making Social Psychology more Ecological and Ecological Psychology more Social". *Ecological Psychology* 19:2, 79–84.

Hutto, D. (2008): *Folk Psychological Narratives*. Cambridge, Mass.: MIT Press.

Hutto, D. D.; Myin, E. (2013): *Radicalizing Enactivism: Basic Minds without Content*. Cambridge, MA.: MIT Press.

Hutto, D. D.; Myin, E. (2017): *Evolving Enactivism: Basic Minds Meet Content*. Cambridge, MA.: MIT Press.

Hutto, D. D.; Satne, G. (2015): "The Natural Origins of Content". *Philosophia* 43:3, 521–536.

Johansson, G. (1973): "Visual Perception of Biological Motion and a Model for its Analysis". *Perception* 6:14, 201–211.

Kiverstein, J. (2015): "Empathy and the Responsiveness to Social Affordances". *Consciousness and Cognition* 36, 532–542.

Knoblich, G.; Butterfill, S.; Sebanz, N. (2011): "Psychological Research on Joint Action". In: B. H. Ross (ed.). *The Psychology of Learning and Motivation*. Burlington: Academic Press, 59–101.

Krueger, J. (2011): "Extended Cognition and the Space of Social Interaction". *Consciousness and Cognition* 20:3, 643–657.

Krueger, J. (2013): "Ontogenesis of the Socially Extended Mind". *Cognitive Systems Research* 25–6, 40–46.

Mark, L. S. (2007). Perceiving the Actions of Other People. *Ecological Psychology*, 19(2): 107–136

Marsh, K. L.; Richardson, M. J.; Baron, R. M.; Schmidt, R. C. (2006): "Contrasting Approaches to Perceiving and Acting with Others". *Ecological Psychology* 18:1, 1–38.

Marsh, K. L.; Richardson, M. J.; Schmidt, R. C. (2009a): "Social Connection through Joint Action and Interpersonal Coordination". *Topics in Cognitive Science* 1:2, 320–339.

Marsh, K. L.; Johnston, L.; Richardson, M. J.; Schmidt, R. C. (2009b): "Toward a Radically Embodied, Embedded Social Psychology". *European Journal of Social Psychology* 39:7, 1217–1225.

Menary, R. (2020). Growing Minds: Pragmatic Habits and Enculturation. In F. Caruana; I. Testa (eds.): *Habits: Pragmatist Approaches from Cognitive Science, Neuroscience, and Social Theory*. Cambridge: Cambridge University Press, 297–319.

Michael, J.; Christensen, W.; Overgaard, S. (2014): "Mindreading as Social Expertise". *Synthese* 191:5, 1–24.

Moreira de Carvalho, E. (2020): "Social Affordances". In: Vonk J.; Shackelford T. (eds.) *Encyclopedia of Animal Cognition and Behavior*. Cham: Springer, https://doi.org/10.1007/978-3-319-47829-6_1870-1.

Pacherie, E. (2013): "Intentional Joint Agency: Shared Intentions Life". *Synthese* 190:10, 1817–1839.

Reddy, V. (2015): "Self in Culture: Early Development". In: *International Encyclopedia of the Social & Behavioral Sciences*, 439–444, DOI:10.1016/B978-0-08-097086-8.23010-9

Reddy, V. (2019): "Why Engagement? A Second-Person Take on Social Cognition". In: A. Newen; L. de Bruin; S. Gallagher (eds.): *The Oxford Handbook of 4E cognition*. Oxford: Oxford University Press, 433–452.

Reddy, V.; Markova, G.; Wallot, S. (2013): "Anticipatory Adjustments to Being Picked Up in Infancy". *PLoS One* 8:2, 195–212.

Reed, E. (1996): *Encountering the World: Toward an Ecological Psychology*. Oxford: Oxford University Press.

Reed, E. S. (1993): "The Intention to Use a Specific Affordance: A Conceptual Framework for Psychology". In: R. H. Wozniak; K. W. Fischer (eds.): *Development in Context. Acting and Thinking in Specific Environments*. New York, NY: Psychology Press, 45–76.

Richardson, M. J.; Marsh, K. L.; Isenhower, R. W.; Goodman, J. R. L.; Schmidt, R. C. (2007): "Rocking Together: Dynamics of Intentional and Unintentional Interpersonal Coordination". *Human Movement Science* 26:6, 867–891.

Rietveld, E.; Kiverstein, J. (2014): "A Rich Landscape of Affordances". *Ecological Psychology* 26:4, 325–352.

Rochat, Ph. (2012): "Early Embodied Subjectivity and Inter-Subjectivity". In: N. E. Coelho; Jr. P. Salem; P. Klautau (eds.): Dimensões da Intersubjetividade, São Paulo: Escuta, 149–164.

Rochat, Ph. (2015): "Self-Conscious Roots of Human Normativity". *Phenomenology and the Cognitive Sciences* 14:4, 741–753.

Runeson, S. (1985): "Perceiving People through their Movements". In: B. D. Kirkcaldy (ed.): *Individual Differences in Movement*. Lancaster: MTP Press, 43–66.

Runeson, S.; Frykholm, G. (1983): "Kinematic Specification of Dynamics as an Informational Basis for Person-and-Action Perception: Expectation, Gender Recognition, and Deceptive Intention". *Journal of Experimental Psychology* 112, 585–615.

Satne, G. (2014): "Interaction and Self-Correction". *Frontiers in Psychology* 5, 1–11.

Satne, G. (2015): "The Social Roots of Normativity". *Phenomenology and the Cognitive Sciences* 14:4, 673–682.

Satne, G. (2016): "A Two-Step Theory of the Evolution of Human Thinking: Joint and (various) Collective Forms of Intentionality". *Journal of Social Ontology* 2:1, 105–116.

Satne, G. (2021): "Understanding Others by Doing Things Together: An Enactive Account". *Synthese* 198, 507–528.

Satne, G.; Salice, A. (2020): "Shared Intentionality and the Cooperative Evolutionary Hypothesis". In: A. Fiebich (ed.): *Minimal Cooperation and Shared Agency*, New York, NY: Springer, 71–92.

Segundo-Ortin, M.; Heras-Escribano, M.; Raja, V. (2019): "Ecological Psychology is Radical Enough: A Reply to Radical Enactivists". *Philosophical Psychology* 32:7, 1001–1023. https://doi.org/10.1080/09515089.2019.1668238

Segundo-Ortin, M. (2020): "Agency from a Radical Embodied Standpoint: An Ecological-Enactive Proposal". *Frontiers in Psychology* 11, 1319. https://doi.org/10.3389/fpsyg.2020.01319

Segundo-Ortin, M., Heras-Escribano, M. (2021). "Neither Mindful nor Mindless, but Minded: Habits, Ecological Psychology, and Skilled Performance". *Synthese*. https://doi-org.proxy.library.uu.nl/10.1007/s11229-021-03238-w

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). "Vervet monkey alarm calls: Semantic communication in a free-ranging primate". *Animal Behaviour*, 28(4), 1070–1094. https://doi.org/10.1016/S0003-3472(80)80097-2

Tollefsen, D. (2005): "Let's Pretend! Children and Joint Action". *Philosophy of the Social Sciences* 35, 75–97.

Tomasello, M. (1999): "Emulation Learning and Cultural Learning". *Behavioral and Brain Sciences* 21:5, 703–704.

Tomasello, M. (2000): *The Cultural Origins of Human Cognition*. Harvard: Harvard University Press.

Tomasello, M. (2014): *A Natural History of Human Thinking*. Harvard: Harvard University Press.

Tomasello, M. (2019): *Becoming Human: A Theory of Ontogeny*. Harvard: Harvard University Press.

Tomasello, M.; Carpenter, M.; Call, J.; Behne, T.; Moll, H. (2005): "Understanding and Sharing Intentions: The Origins of Cultural Cognition". *Behavioral and Brain Sciences* 28:5, 675–691.

Trevarthen, C. (1979): "Communication and Cooperation in Early Infancy: A Description of Primary Intersubjectivity". In: M. Bullowa (ed.): *Before Speech: The Beginning of Human Communication*. Cambridge: Cambridge University Press, 321–347.

Vaish, A.; Missana, M.; Tomasello, M. (2011): "Three-Year-Old Children Intervene in Third-Party Moral Transgressions". *British Journal of Developmental Psychology* 29:1, 124–30.

Wagman, J. B. (2019): "A Guided Tour of Gibson's Theory of Affordances". In; J. B. Wagman; J. C. Blau (eds.): *Perception as Information Detection. Reflections on Gibson's Ecological Approach to Visual Perception*. New York, NY: Routledge, 130–148.

Short Biography

Miguel Segundo-Ortin, PhD, is a post-doctoral researcher at Utrecht University (The Netherlands). He specializes in the philosophy of the cognitive sciences, with a special emphasis on embodied and situated theories of cognition. Other research interests include the study of so-called 'minimally cognitive agents', the explanation of skilled performance, and the relationship between culture and cognitive development. Currently, Miguel is working on a research project that aims to develop an embodied and situated account of human self-control.

Glenda Satne, PhD, is a Senior Lecturer in Philosophy at the School of Liberal Arts at the University of Wollongong, Australia. She works on Philosophy of Mind and Social Ontology and has published extensively on Embodied cognition, Enactivism, Collective Intentionality, the Second-Person, Normativity, and Cultural Evolution.