The public character of visual objects: shape perception, joint attention, and standpoint transcendence

Axel Seemann[1]

Abstract

Ordinary human perceivers know that visual objects are perceivable from standpoints other than their own. The aim of this paper is to provide an explanation of how perceptual experience equips perceivers with this knowledge. I approach the task by discussing a variety of action-based theories of perception. Some of these theories maintain that standpoint transcendence is required for shape perception. I argue that this standpoint transcendence must take place in the phenomenal present and that it can be explained in terms of the experience of perceivers who jointly attend to an object. Joint perceivers experience objects as being perceived from standpoints other than their own. They operate in what I call "social space", in which they single out objects by triangulating targets' locations relative to their co-perceivers' standpoints on these targets. It is then possible to explain the public character of the objects of individual experience by appeal to what I call "public space". This is a spatial framework whose locations are presented as standpoints whence joint attention to the target would ensue, were they occupied by co-perceivers. If shape perception requires standpoint transcendence, then shape perceivers operate in public space and are thus capable of singling out targets by triangulating their locations from standpoints other than their own. If it doesn't, then the introduction of a public spatial framework is an additional step whose introduction explains how perceivers come to experience objects as perceivable from standpoints other than their own.

✉ Axel Seemann
aseemann@bentley.edu

[1] Department of Philosophy, Bentley University, 175 Forest Street, Waltham, MA 02452, USA

Springer

1 Introduction

We live in a world in which the ordinary objects of visual experience[1] can be seen by other perceivers also and in which we know that they are perceivable from standpoints other than our own. The question I address in this paper is how this knowledge is possible. The answer I offer takes as its premise that the knowledge in question is of an experiential kind: it is because of the way in which visual objects are presented in experience[2] that we know they are perceivable, concurrently, from other standpoints also. The challenge then is to explain how experience can deliver this feat: how can it be that visual experience, which presents its objects relative to the standpoint occupied by the perceiver, nevertheless equips the perceiver with the knowledge that they are perceivable from standpoints other than her own?

My proposal is, in short, that perceivers know that objects are perceivable from standpoints other than their own because ordinary experience presents locations in visual space as standpoints whence, were they occupied by co-perceivers, the object would be jointly perceived.[3] Human perceivers' capacity for joint attention explains how they know that the objects of visual experience are public. I substantiate this claim by appeal to the hypothesis of social space, according to which joint attention is facilitated by a spatial framework in which the locations of perceiver and co-perceiver are presented as standpoints, so that the perceiver can single out a target object by triangulating its location relative to the co-perceiver's standpoint. I then introduce the hypothesis of public space, according to which ordinary individual perceivers operate with a spatial framework in which locations not currently occupied by co-perceivers are presented as standpoints whence joint attention to a target would ensue, were they so occupied. When philosophers think about joint attention, they tend to explain it, "bottom up", by building on individual perception. The hypothesis of public space turns this explanatory order upside down: we can know in experience that the objects of perception transcend our viewpoints on them because we operate with a spatial frame of reference in which we could always point them out to others and thus come to jointly attend to them with these others. Operators

---

[1] By "ordinary objects" I mean three-dimensionally extended visible space occupiers. Thus, your sofa is an ordinary object but a hallucination of your sofa is not.

[2] Throughout this paper, I take perceptual objects to be "presented" rather than "represented" in experience. This is because of my commitment to a social form of epistemological disjunctivism (Seemann, 2019, pp. 67–72), the view that only joint experiences justify common knowledge claims about their objects; and because this form of disjunctivism is not obviously compatible with representationalism about experience.

[3] A qualifier is in order: the object would be jointly perceived if currently unoccupied standpoints in a perceiver's visual space were occupied by a co-perceiver and if a range of conditions are met. These include the absence of a visual barrier between the location of the co-perceiver and that of the target object, the object's having the right size (it neither being too small nor too big to be visible in its entirety), the perceiver's eyesight being sufficient, amongst others. Throughout this paper, I assume that these conditions are met.

in public space are always potential joint perceivers, and this explains the public character of the way in which ordinary objects are presented in individual experience.[4]

I approach the question of phenomenal standpoint transcendence by way of a discussion of shape perception. Objects that have shapes are three-dimensional space occupiers. Since we experience objects as having shapes even though we cannot perceive those parts of an object's surface that are hidden from view, it is instructive to consider theories of shape perception in the attempt to explain how standpoints are transcended in experience. After some terminological housekeeping, I offer a brief sketch of an argument in defence of the view that our knowledge of the standpoint transcendence of ordinary objects is to be explained by appeal to experience (2). I then turn to the question of whether certain action-based theories of vision have the resources to account for the public character of the experience of ordinary visual objects. I consider Grush's (2001, 2004, 2007) skill theory of vision as an account that explains shape perception without appeal to the perceiver's ability to transcend her own standpoint and Noe's (2004, 2005) sensorimotor theory as an account on which shape perception requires standpoint transcendence. I conclude that neither account can, by itself, explain the phenomenal presentation of ordinary objects as public (3). In a next step, I describe joint attention as a social form of perception that plays out in a spatial order in which at least two locations are presented as standpoints to each perceiver, and in which perceivers each single out the target by socially triangulating its location relative to the standpoint of their co-perceiver. This social triangulation is to be thought of as a basic kind of deictic social activity that plays out in the phenomenal present (4). I then build on the notion of standpoint transcendence in the phenomenal present to introduce the notion of "public space". In public space, locations other than the perceiver's own are presented as standpoints whence joint attention to a target object would ensue, were they occupied by co-perceivers. I develop this view by discussing two further action-based theories of shape perception, Schellenberg's (2007) notion of an "alter-ego-centric" conception of space and Kelly's (2004) interpretation of Merleau-Ponty's (1945/2002) contention that the "real object" is seen "from everywhere". I suggest that visual perceivers who experience objects as public operate with a practical conception of space in which background locations are standpoints relative to which the perceiver would socially triangulate the location of the target, were they occupied by co-perceivers (5). I end with some concluding remarks (6).

---

[4] I should highlight the somewhat speculative nature of this proposal. The notions of social space and of public space are hypotheses that are in need of further substantiation (though see Seemann (2019, chs. 10&11) for an extended discussion of social space). The viability of the present proposal depends on whether these hypotheses can be defended.

2

I begin with some terminology. A "standpoint" is a location in allocentric space[5] that, when occupied by a perceiver and agent, serves as an origin of perception and action relative to some target object.[6] A standpoint is "transcended" in experience if the experience presents a target object as perceived, or perceivable, from standpoints other than the perceiver's own.[7] Standpoint transcendence is "individual" when achieved by a single perceiver and "social" when achieved in joint attention. "Joint attention" is a social mode of object perception that involves at least two perceivers and in which each perceiver experiences a target object as being singled out from both her own standpoint and that of the co-perceiver, or co-perceivers.[8] A "co-perceiver" is a perceiver who jointly attends to an object with another perceiver.[9] Visual objects and their properties are "public" if they can be apprehended in experience by subjects other than the perceiver. Experiences in which a perceiver transcends her standpoint present their objects as public. The "publicity requirement" is the need of an explanation of how individual perceptual experiences can present their objects as public.

I take it, without further discussion, that human perceivers and agents know that the ordinary objects of visual experience are perceivable by other perceivers also. That they should be so perceivable is a fundamental part of our everyday metaphysics. In ordinary circumstances we simply assume that I can point out the objects I see to you, so as to facilitate shared action, transmit perceptual facts and produce perceptual common knowledge about them. This is true even in the absence of other perceivers. Then the question

---

[5] My use of the notion of "allocentric space" follows Grush (2001, p. 80): it is a space that "has another object, person, or perhaps just location as its origin."

[6] A reviewer wondered what the difference was between a centre of action and perception in egocentric space and a standpoint. An egocentre is an origin of perception and action—a point in space relative to which the location of various objects can be described indexically ("to my left", "to my right", etc.); see Section 8. It need not itself be specified relative to other objects in allocentric space. By contrast, a standpoint is a location that can be specified relative to other objects and happens to be occupied by a perceiver. This location therefore also serves as but is nevertheless not identical to an egocentre, You cannot occupy an egocentre other than your own, but you can change your standpoint on a target object by moving relative to it. This difference matters because it follows that standpoints but not egocentres can be transcended.

[7] There are at least two ways in which standpoints can be transcended. First, you can transcend your standpoint on a perceptual object individually, by moving around it while keeping track of it or by encountering it for a second time while remembering that you have seen it from a different standpoint before. Secondly, and as I shall argue, you can also transcend your standpoint by jointly attending to an object with another perceiver.

[8] This interpretation of the notion of joint attention is not neutral relative to all possible accounts. For instance, on a "lean" account of joint attention (Racine, 2011) there is no experiential dimension to joint attention. I follow Campbell (2002, 2011) in treating joint attention as an experiential phenomenon. Discussion is not possible here.

[9] I sometimes say, in the interest of brevity, that in public space locations are standpoints whence joint attention to a target would ensue, were they occupied by co-perceivers. Given my definition of co-perceivers as perceivers with whom one jointly attends to an object, this description is, strictly speaking, circular. What I mean by it is this: locations in public space are standpoints whence, if they are occupied by other perceivers who enter in a deictic communication about a target object with the perceiver and if the conditions laid out in ft. 3 are met, joint attention to the target will ensue.

arises how we can know that the objects of experience are public. There are at least two different ways in which we could gain this knowledge. One possibility is that you know that an object is perceivable from standpoints other than your own present standpoint because you have encountered it from other standpoints in the past, or because you keep track of it as you move around it. But we usually know that the objects of perception are public even if we have never encountered them before or if we see them without any change in spatial relation between ourselves and these objects. A different explanation is that this knowledge is produced by an inference from objects' three-dimensional appearance. On some views (see the next section), the three-dimensional appearance of visual objects can be explained without appeal to standpoint transcendence. Then the perceiver might infer from the experience of the object as having depth that it is visible from standpoints whence visual information from the object's currently hidden aspect would be available. Such an inferential account has certain advantages. For example, it can explain the fallibility of perceivers' knowledge of visual objects' public existence.[10] But it faces significant problems also. First, the inference could not be of a conscious kind: it is not that you see a three-dimensional object and then work out that it is public. Secondly, it is not obvious that the experience of an object as three-dimensional enables a perceiver to infer its publicity unless the perceiver is already operating with the concept of a standpoint. Then you need to explain how the perceiver can have this concept. And an attractive way to deliver this explanation is to say that the perceiver has the concept of a standpoint because she has learned to distinguish between her own and other possible standpoints on some object. But if she has understood this distinction, she has already mastered the concept of publicity, since a public object just is one on which different standpoints are possible. So the explanation is circular.

I am not suggesting that these cursory remarks show that an inferential account of knowledge of visual objects' publicity is in principle impossible; it is just that such an account faces considerable difficulties. To keep this paper at manageable length, I bracket the possibility of alternative accounts of knowledge of publicity. I develop an answer to the question of whether this knowledge is explicable by appeal to the way in which public objects are presented in experience, but I do not claim that my proposal rules out other, non-experiential approaches.

3

In his critical evaluation of Noe's sensorimotor theory of perception, Briscoe (2008, pp. 491–492) approvingly quotes a remark of Merleau-Ponty's about the equivocality of the perspectival character of knowledge:

"It can signify that only the perspectival projection of objects would be given to primitive knowledge; and in this sense the expression is inexact since the

---

[10] It can, on some views at least, in principle turn out that what looks to be your public sofa is your private hallucination of the sofa. See the essays collected in Macpherson and Platchias (2013) for discussion.

first reactions of an infant are adapted, e.g., to the distance of objects—a fact that excludes the idea of a phenomenal world originally without depth.

At the same time, the perspectival character of knowledge provides perception

... with the assurance of communicating with a world that is richer than what we know of it, that is, of communicating with a real world (1963, pp. 186-187)."

There are, on this view, two questions that arise with regard to the perspective-dependence of perceptual knowledge. One question is how objects can be presented in experience as being extended in three-dimensional space, and thus as being distant from the perceiver as well as having shapes. Another question is how objects can be presented as outstripping the perceiver's perspective, in the sense of having aspects that are hidden from the perceiver's view but are visible from somewhere else. In this section, I sketch two theories that take up these questions in different ways. The first one, Noe's (2004) sensorimotor theory, begins with the assumption that primitive vision is two-dimensional and that an additional step is required to explain how we can see objects as being positioned and extended in three-dimensional space, and thus as having shapes. On such a view, Merleau-Ponty's two questions are answered together: to explain an object's shape you have to think of it as an object whose surface is always partly hidden from view and that is thus "real" in the sense of the above quote. The second theory is Grush's (2001, 2004, 2007) skill-based account, on which objects are presented as three-dimensional from the outset and on which this can be explained without appeal to standpoint transcendence, and thus without having to consider questions about their perceivability from standpoints other than the perceiver's. This account addresses the first of Merleau-Ponty's questions without also providing an answer to the second.

Both are members of the family of action-based accounts of perception – theories, that is, that seek to explain our ability to perceive objects' shapes by way of the perceiver's and agent's exercised or dispositional knowledge of how one has to move in order to act on such objects or how their appearance would change, were one to move around them.[11] I focus on this family of theories because I shall be seeking to explain perceivers' knowledge of objects' publicity by appeal to the standpoint-transcendence inherent in joint forms of perception, and because I think of joint attention as a kind of activity. I am making no claim that only action-based theories of vision stand a chance of success, and I am not claiming that the two accounts sketched below are the only plausible candidates for such a theory. My aim here is not to adjudicate between the candidates' promise for an explanation of shape perception but to assess their usefulness for an account of the public character of the objects of visual experience.

3.1 The sensorimotor theory

The sensorimotor theory has its roots in the sense data theory of perception (Briscoe, 2008). On this view, visual objects are presented as two-dimensional patches on a

---

[11] See Briscoe and Grush (2020) for an overview of action-based theories of perception.

flat surface. An extra step is then required to explain how we come to grasp them as having three-dimensional shapes. On Noe's (2004) account, shape perception is to be explained in terms of an implicit knowledge of how an object's appearance would change, were you to move around it. Seeing an object as having a specific shape requires both receiving sensory stimulation from the object and using those stimulations to retrieve the set of sensorimotor contingencies associated with the object on the basis of past encounters. Hence there are two steps to seeing an object as having a shape: the perceiver experiences the object as a two-dimensional patch (its "perspectival shape") that would occlude the object on a plane perpendicular to the line of sight; and the perceiver understands how the patch's appearance would vary with changes in her point of view. On this view, perceiving is an activity, a way of skillfully exploring the world (Noe, 2005, p. 244). Experience of an object's shape is given only to the extent that the perceiver enjoys implicit practical knowledge of changes in the object's sensorimotor profile, depending on how the perceiver's standpoint changes. So the perceiver is relying on standpoints that are spatially distinct from her present one and that she has occupied in the past to retrieve the object's sensorimotor profile. Experiencing an object as being three-dimensionally extended thus requires a form of standpoint transcendence.

The account has received its fair share of criticism. Briscoe (2008) marshals an impressive amount of evidence against the view that an object's visually apparent shape is two-dimensional. Schellenberg (2007, pp. 609–611) asks how the account can explain the experience of objects whose shape type we have not encountered before. She also argues that, insofar as the account requires two encounters with an object, and given that the object's appearance may be different in each encounter, it is not clear how one can tell that the two encounters are with the same object. Furthermore, it is not clear how these distinct appearances are integrated into the perception of a single object. These objections are vital for the present project: if a perceiver is to know, on the basis of experience, that an object is perceivable from standpoints other than the perceiver's own, there has to be an explanation of how the perceiver can know that the occupant of these standpoints is experiencing the same object as the perceiver herself. If Schellenberg is right and the sensorimotor theory does not deliver such an explanation, it is not a suitable starting point for an account for the public character of the objects of visual experience. The problem she identifies for the sensorimotor theory as a theory of shape perception is bound to recur for such an account, since it just is a theory of how standpoints on particular visual objects are transcended in experience.

3.2 The skill theory

The second theory explains shape perception without appeal to standpoint transcendence. It has a forerunner in Gareth Evans's work on spatial representation. In an oft-quoted passage, he writes:

> Egocentric spatial terms are the terms in which the content of our spatial experiences would be formulated, and those in which our immediate behavioural plans would be expressed. This duality is no coincidence: an egocentric space can exist only for an animal in which a complex network of connections exists

between perceptual input and behavioural output. A perceptual input . . . cannot have a spatial significance for an organism except in so far as it has a place in such a complex network of input-output connections. (Evans, 1982, p. 154)

The connection between spatial representation and shape perception is as follows: if a creature is to perceive an object's shape, it has to perceive the thing as being extended in space. For that to be possible, the creature has to be operating with a spatial framework that enables it to perceive the object's extension. For Evans, the vital consideration is that objects are presented in egocentric space. That is, they are presented in a framework in which objects are placed so that their location could be described in indexical terms ("to my left"; "to my right", etc.). But a creature need not be capable of so describing the location of perceptual objects; it only needs to know how to act on them. Egocentric space is constituted by this practical know-how. You can then explain shape perception in terms of the spatial framework that facilitates action on the perceived object. You see an object's shape because you know how you would have to move in order to act on the thing. This proposal is importantly different from the sensorimotor account. It does not explain shape perception in terms of expectations about sensorimotor information that will become available from a type of object if the perceiver moves relative to it. It appeals to agency: it explains shape perception in terms of the practical knowledge of what is required to purposefully manipulate a particular thing.

Grush's skill theory of perception develops this view. I can only give a sketch of some relevant aspects of this complex and sophisticated theory here. The core idea, for present purposes, is that having a point of view is tantamount to operating within egocentric space; the point of view itself is egocentric space (Grush, 2001, p. 77). This space is not defined by what is currently sensed but rather by the actions that it makes possible. Shape perception is then not explained by appeal to the perceiver's ability to transcend his own standpoint; it is built into the notion of a standpoint itself. Grush (2001, p. 65f.) illustrates his view by considering sensory substitution devices. Such devices translate visual into tactile information, in his example by vibrations on a subject's back. Astoundingly, they enable blind people to experience spatial layouts in a way that allows them to act on objects in their vicinity. It is not that they learn to infer from a particular kind of signal where an object is located ("If there's a slow vibration near my left shoulder then there's an object within reach of my left arm"); they directly sense the object's place ("There's an object over there"), much in the way in which the thing would be presented in visual experience. Grush uses this finding to make vivid a point that is at the heart of Evans's thinking about egocentric space: vision is not inherently spatial (Grush, 2001, p. 66); awareness of the spatial order of the visual environment is not the foundation but the result of an agent's skill to act on its objects. In the roughest possible form, such skills consist in functions that connect sensory input with dispositions to act. Grush calls the relevant kind of disposition "detail specifying": it "is a disposition that, for any given behavior type (such as a grasp or a foveation, or whatever) specifies the details of how this behavior type will be executed, if it is executed" (Grush, 2007, p. 393). Since functions from sensory input to dispositional output do not by themselves enable a subject to predict the consequences of any of their movements, and since the capacity to act depends

on such predictions, Grush (2004) relies on emulation theory to explain perceivers' capacity to predict the consequences of movement. The main idea is that during sensorimotor engagement the brain models the agent in the environment. These models are used to process sensory information and to generate expectations of sensory feedback. They can also be used, off-line, to generate expectations about outcomes of different actions and to develop motor plans. Shape perception can then be explained by appeal to "pattern concepts" (Grush, 2007, p. 407). The agent who operates in egocentric space, and who thus both is in a position to know how she has to move in order to act on the objects it contains and to predict what the consequences of her sensorimotor engagement will be, can form expectations about the spatial extension of perceptual objects. Possession of a pattern concept of a certain spatial arrangement allows the perceiver to "fill in" those parts of the representation of the complete object that are not directly provided by sensory information.

This theory understands visual object perception as three-dimensional from the outset, without appeal to the kind of standpoint transcendence required on Noe's view. So the account does not by itself explain how the publicity requirement, with the standpoint transcendence it entails, can be met. One possible such explanation is that the agent who knows that ordinary visual objects are public is operating with a conception of space. A perceiver who enjoys a conception of space can think of herself as occupying locations that are distinct from the standpoint she is occupying (Evans, 1982, 163). She is then able to work out, for instance, how she would have to move in order to get to a destination from a location other than her present one. She has transcended the egocentric order of the region in which she is capable of action. But the notion of a conception of space is not the right concept for explaining how the publicity requirement is met in perceptual experience. This concept is not phenomenal. It does not, by itself, say anything about the character of the experience of the creature who operates with it. It is, also, developmentally unduly demanding. Someone who points out objects to others knows that the target of her pointing gesture is visible to others also; she is operating with a conception of the object as public. One-year old children begin to undertake such acts of pointing. But it is not credible that they are operating with a full-blown conception of space in Evans's sense.[12]

In this section, I have looked at two ways of explaining shape perception. One of them makes use of the idea that seeing an object's shape requires previous encounters with an object of the same shape-type. The perspective transcendence that makes shape perception possible is then diachronic and therefore not a promising starting point for an account of phenomenal publicity. The second proposal does not appeal to standpoint transcendence at all and for that reason does not, on its own, have the resources to explain the public character of visual objects. What is needed is an account that can explain how standpoints are transcended in

---

[12] Having a conception of space in Evans's (1982, p. 162/163) sense requires that one be able to locate one's egocentre on a cognitive map, so as to be able to generate counterfactual hypotheses about what one would observe from locations one is not currently occupying. Though the onset of counterfactual thinking in children is debated, the earliest evidence for implicit reference to counterfactuals is at 2.5 years of age (Beck et al., 2011) and thus significantly later than the early stages of joint attention.

the phenomenal present,[13] without requiring previous exposure to the same object or type of object. Objects that are presented, in the phenomenal present, as being perceived from standpoints other than the perceiver's own are necessarily public. In the next section, I consider joint attention as a form of perceptual experience in which the publicity requirement is met and then argue that we can build on the joint case to explain the public character of individually perceived objects.

4

I now explain how standpoints are transcended in joint attention. The core view is that joint perceivers experience target objects as being singled out relative to standpoints other than their own in the phenomenal present and thus come to know that they are being perceived by the occupants of these standpoints.[14] In joint attention, target objects are presented as public because perceivers transcend their own standpoints on the target. For this to be possible, the co-perceiver's focus on the object has to feature in the perceiver's experience. It is notoriously difficult to make this requirement precise. The first thing to note is that I am concerned with the phenomenology rather than the metaphysics of joint experiences. So I am not concerned with the nature of the perceptual relation that underwrites joint attention. I only seek to bring out those aspects of its phenomenology that explain how joint perceivers transcend their own standpoints in experience. The account I shall be introducing is not neutral relative to all possible views on joint attention,[15] and it is also not neutral relative to all possible ways of conceptualizing the triadic relation that obtains between joint perceivers and their object of attention. But I develop it by appeal to considerations about how joint attention is initiated and maintained that lend at least some empirical support to the action-based account of joint attention I am putting forward.

It seems uncontroversial that joint attention requires continued bodily activity by the perceivers for its initiation and maintenance. It involves pointing, gaze-following, and other purposive or non-purposive movements that support the perceiver's focus on a particular target and the co-perceiver's interpretation of those movements as signifying the perceiver's focus on the target (e.g., Mundy et al., 2007). For this to be possible, the perceivers must be able to work out where the object is placed relative to the co-perceiver's location. The object of joint attention is a space occupier that the perceivers each know to locate, by triangulation, relative to their co-perceiver's standpoints. Consider how an episode of joint attention is brought about and maintained. In joint attention, each perceiver alternates their focus between the

---

[13] By "phenomenal present" I mean an experience that the perceiver would describe as occurring "now", in the present. I say more about this notion in Section 4.

[14] As already noted, this experiential view of joint attention is not the only possible one. I adopt it here without further discussion. For an extended discussion that includes the connection between joint attention, experience, and shared forms of practical and theoretical knowledge, see Seemann (2019).

[15] See the papers collected in Eilan et al. (2005) and Seemann (2011b) for an overview.

target object and the co-perceiver, so that the perceiver can find out which object[16] the other perceiver is making salient to her, and so as to make that object salient to the co-perceiver.

This kind of triangulation is unlike the procedure by means of which a perceiver can locate an object in allocentric space (Grush, 2001, pp. 79–80) relative to the position of another object. This procedure is mastered by many animals incapable of joint attention. Perceivers who are capable of what I call social triangulation single out a target object relative to the location of a co-perceiver and thus relative to a standpoint. One way in which they could accomplish this is by having solved the problem of other minds: if a perceiver knows that a certain physical object is a perceiver and agent, she thereby also knows that the location it occupies is a standpoint. But it is not obvious that joint perceivers always enjoy theoretical knowledge of other minds. There are various views on what it takes to acquire this knowledge,[17] but none of them suggests that it is available around the end of the first year of life, when the capacity for joint attention sets in. This might be seen as spelling bad news for the view that children who have not solved the problem of other minds are capable of genuine forms of joint perception.[18] To solve the problem, I introduce the hypothesis of "social space". It suggests that creatures who jointly attend to a target operate with a spatial framework in which the locations occupied by co-perceivers and co-agents are presented as origins of perception and action and thus as standpoints. On this view, it is the spatial framework of the perceivers, not (or at least not necessarily) their knowledge about its occupants, that explains how a location can be presented as a standpoint and how social triangulation is thus possible.

Standpoints are origins of perception and action. Action requires that the agent integrate sensory information from a variety of sources at the location she is occupying.[19] My vision guides the movement of my hand with which I reach for the cup; it enables me to move my arm in its direction and, once I have reached it, aids the movement of my fingers in closing my hand round it. But my fingers are also guided by my sense of touch, and my entire movement is guided by proprioception, the internal sense that tells me about the movement of my limbs and their relative position to each other. Action relies on the multisensory integration of information in the area in which object-directed motor movement takes place. The hypothesis of social space builds on the notion of action space to suggest that in joint attention, agents use their own bodily

---

[16] The locution "which object" is imprecise. For present purposes, the object in question is the thing at the location identified by the participants in an episode of joint attention through social triangulation. For a treatment of the question how this definition deals with distinct objects occupying the same location and distinct objects overlapping at one location, see Seemann (2019, ch. 5).

[17] See Avramides (2001) for an overview.

[18] One possibility is that infants can directly perceive others as minded in social interaction (e.g., Gallagher, 2008). As a reviewer of this paper pointed out, it is of course also possible that full-fledged, adult joint attention is phenomenally and cognitively quite different from the early triadic interactions that one-year olds begin to engage in. Discussion is beyond the scope of this paper.

[19] See Rizzolatti et al. (1997) for a seminal paper that argues that there is a specific kind of spatial map represented in the brain that is responsible for motor movement. De Vignemont (2018) discusses the relation between egocentric space and action space.

resources to present other agents as co-perceivers and their locations as standpoints. In social space, perceivers integrate sensory information in a variety of modes both at their own location and the location of their co-perceiver.[20] More particularly, they bind proprioception from their own bodies with visual information from the co-perceiver's body, so that a direct, non-inferential awareness of the other's bodily movement as a motor activity is attained and the other's location is presented as an action space. These action spaces are, in a sense, impoverished: you cannot have direct access to the other's sense of touch, you cannot have visual information about objects from perspectives other than your own, you do not enjoy a sense of control of the other's movements, and the presentation of a location not occupied by you as a standpoint appears contingent on your own location being presented as a much more fully developed, richer peripersonal space. But the integration of bodily with external sensory information at the co-perceiver's location suffices to establish a spatial framework in which others' locations are presented as standpoints to each perceiver and in which these joint perceivers therefore transcend their own standpoints. Standpoints are transcended in social space because joint perceivers and agents use internal sensory information from their own bodies to present an object at a given location as a co-perceiver, which thus becomes an origin of motor action.

The hypothesis of social space is of a psychophysiological kind. But it makes available an account of the phenomenal dimension of joint attention. This dimension consists in the presentation of the other's extended location as an action space at which the perceiver integrates sensory information in a variety of modes. You experience other agents as co-perceivers because you use internal sensory information from your body to apprehend their movements as actions. Because in joint attention the target object is singled out, by each perceiver, relative to the co-perceiver's standpoint, and because there is a phenomenal dimension to the presentation of that standpoint, joint perceivers come to know in experience that the target of their attention is being perceived from a standpoint other than their own and that it is hence public.[21]

An important question arises about the temporal dimension of episodes of joint attention. On the approach I have been presenting, joint attention is a process that is extended in time. But you can at least sometimes come to perceptually know that an object is being jointly perceived, and thus that it is being seen from standpoints other than your own, in the phenomenal present. I am not denying that there are occasions, such as my laborious efforts to make a particular bird that is sitting on a remote tree branch salient to you, that require sustained activity over time. But not

---

[20] See Seemann (2019), ch. 11, for a discussion of the empirical findings that support the hypothesis of social space.

[21] I am not suggesting that this account exhausts the phenomenology of joint attention. One important family of views suggests that it is the sharing, or attunement, of feelings or emotions between perceivers that distinguish the experience of joint attention from other forms of object perception (e.g., Hobson & Hobson, 2005; Seemann, 2011a; Trevarthen, 1980). The current proposal does not deny the relevance of intersubjectivity for a complete account of the phenomenology of joint attention. I am concerned here only with spatial awareness in joint attention, since it is that aspect of the experience of joint perceivers that can help explain the public character of the objects of shared attention.

all joint attention is like that: it is often established at a glance, in a single look. You can come to know perceptually that someone is pointing out a target to you in the phenomenal present, particularly if the target is clearly visible and positioned close to your co-perceiver. The challenge is to explain how this is possible. Meeting the challenge is vital for the present proposal: joint attention is argued to offer a solution to the problem of standpoint transcendence precisely because it plays out in the phenomenal present. If it turns out that standpoints are transcended over time in shared forms of perception, then the account fails to solve the problem.

Much will depend here on how the notion of the phenomenal present is conceptualized. By "phenomenal present" I mean an experience that the perceiver would describe as occurring "now", in the present. This does not entail that the experience does not have duration (see Gallagher et al. (2017); Zahavi (2011) for discussion). The phenomenal present is not a freestanding snapshot without any connection to the past and the future. It has a temporal structure and thus a minimal temporal extension. But it nevertheless is presented to the perceiver as occurring "now". The suggestion is that joint perceptual episodes that require no or only minimal attentional and deictic effort play out in the phenomenal present, while those that are effortful take place over a phenomenally extended period of time. In joint episodes that are experienced as occurring now, a co-perceiver's pointing gesture very quickly leads the perceiver's focus to a given target. Endogenous, intentionally directed attention takes, at 300 ms, about three times as long to be deployed as its exogenous counterpart (Carrasco, 2011). Furthermore, intentionally directing attention to a stimulus lengthens the subjective time of the episode (Matthews & Meck, 2016). Suppose you are hearing a loud bang and reflexively turn your head in the direction of its origin; or suppose you encounter a big arrow and automatically follow its lead. Your experience is not of first hearing the bang or seeing the arrow and then turning your head. The whole episode, despite having duration and being temporally structured, is presented to you as occurring now. In the same way, an episode of joint attention that makes the intended object of the co-perceiver's deictic gestures salient without requiring significant interpretive effort is experienced as occurring in the present. Nevertheless, there is a fundamental difference between this cotemporaneous form of joint attention and mutual gaze following, in which automatic responses to perceptual cues lead to the same target being perceived without any kind of shared experience. The difference between these two kinds of events is that joint attention plays out in social space, in which the location of the co-perceiver is presented as a standpoint and the target is singled out relative to that standpoint by the perceiver.

5

On the view I have been sketching, the publicity requirement is met in joint attention because joint perceivers experience targets as being perceived, in the phenomenal present, from standpoints other than their own. In this section, I consider two action-based accounts of shape perception that appeal to standpoint transcendence in the phenomenal present. This distinguishes them from both Grush's skill-based theory, on which standpoint transcendence is not required to explain shape perception, and

Noe's sensorimotor theory, on which standpoint transcendence is achieved by previous encounters with objects of the same shape type. I argue that for any theory that seeks to explain shape perception by appeal to standpoint transcendence in the phenomenal present, the locations that serve as standpoints have to be standpoints whence joint perception of the target would ensue, were they occupied by perceivers. For this family of theories, three-dimensional objects are necessarily public.

5.1 Alter-ego-centric space

Schellenberg (2007) introduces the notion of a practical "alter-ego conception of space" in order to explain how objects can be presented as having three-dimensional shapes. Like the views discussed in Section 8, this account explains shape perception by appeal to agency. Differently from these views, it does this in terms of the spatial organization of the environment in which three-dimensional objects are presented. On Schellenberg's view, a creature who is operating with an alter-ego conception of space knows how she would have to move to act on an object, were the spatial relation between herself and the object to change. This practical knowledge is spelled out in terms of the notion of a "capacity to act" (Schellenberg, 2007, p. 622f.) that someone can have even if it is not currently being exercised. The second step is to argue that shape perceivers self-represent their own locations as centres of perception and agency (Schellenberg, 2007, p. 620f.). This self-representation is not relational but "monadic" in the sense of Campbell (1994, p. 119): even though it is made possible by the spatial relation between perceiver and object, it only represents the perceiver's location, whence objects are thus presented as being "on the left" rather than "on my left". It does not involve, or require, self-consciousness. It is a spatial kind of awareness that underwrites the perception of objects in egocentric space. The perceiver treats a variety of locations in allocentric space as "alter-ego vantage points": as viewpoints or centres of agency that, were they occupied by the perceiver, would bring about changes in her disposition to act. Spatial concepts such as "to the left" or "to the right" can then be explained in terms of this dispositional knowledge, and the phenomenal character of the perception of three-dimensional objects can be described as the object's being perceived "as perceivable from different points of view" (Schellenberg, 2007, p. 617).

I now argue that the self-representation of the perceiver relative to the target object, though it need not amount to a substantive kind of self-consciousness, cannot be monadic in Schellenberg's sense. Campbell, in his discussion of the difference between monadic and relational spatial notions, argues that the spatial content of vision can be described in monadic egocentric terms. To describe the location of a visual object, on his view, it is not necessary to invoke the relation between oneself and the thing. Regardless of whether this view is correct for visual spatial relations,[22] it cannot be applied to the conceptualization of alter-ego standpoints. Suppose you see a book on the table before you and you treat the location of the pen to its left as an alter-ego vantage point in Schellenberg's sense. You can then describe

---

[22] See Schwenkler (2014) for an argument to the contrary.

the book's location relative to the pen in two distinct ways. You can describe its location as it would be presented to a perceiver occupying the location of the pen. In a case of ordinary vision, in which the perceiver singles out a target object in attention, the object is then typically presented as being "ahead". This is a monadic notion in Campbell's sense. It is the same notion that the perceiver would use to describe the location relative to her own standpoint. The description could therefore not capture the alter-ego-centric nature of that standpoint; it could not capture the difference between the perceiver's own standpoint and the standpoint a perceiver would have, were she to occupy the location of the pen.

You might think that this is just a consequence of the object's being perceived as being ahead from both locations. But consider a different case: suppose an object is positioned so that it is being perceived by you as being, monadically, on the right and that it would be perceived from a given alter-ego-centric standpoint as being on the left. Now there is a descriptive difference between the location of the object relative to the two locations. The problem is that for this difference to be useful in explaining shape perception, there has to be some factor that guarantees that it is the same object that is being perceived from both locations. What guarantees the sameness of the object is its occupancy of the same location in space. The spatial order of the region in which the object is placed then cannot be egocentric, since in egocentric space there could not be standpoints other than the perceiver's own, relative to which all other locations are describable. But neither could it be allocentric, since the object is then not presented so that the perceiver can enjoy the dispositional knowledge of how to act on it. The sameness of the object would be guaranteed, however, if it were perceived in social space, in which the perceiver treats locations other than her own as standpoints and triangulates the object's location relative to these standpoints. But doing so requires her to invoke relational notions. She could not simply be singling out (say) the alter-ego vantage point as "to the right" (relative to her own standpoint), and then the object as "to the left" (relative to the vantage point), since she would then have no way of knowing that the same object was being singled out by that description as the one she is perceiving herself. What she has to do is identify the vantage point relative to a location she is indexically identifying as "mine" or, perhaps, "here".[23] Then she can treat the vantage point as affording a standpoint on the same object as the one she is presented with from her own point of view.

The general point is that, even supposing you accept that the spatial content of vision can be specified monadically, it is not possible to make sense of the notion of an alter-ego vantage point in terms of monadic visual content. Alter-ego vantage points are not just points in space; they are potential origins of perception on target objects. If they are to be identified so that they afford perspectives on the perceiver's target, relational notions have to be invoked in the triangulation that is required to single out the same thing relative to that vantage point. So the question arises what kind of spatial framework underwrites the notion of an alter-ego vantage point. I

---

[23] The idea that acting in peripersonal space gives rise to a minimal sense of "here" is explored, in different ways, in Seemann (2019, ch. 12) and De Vignemont (2021).

said in Section 8 that a full-blown conception of space in Evans's sense, on which the perceiver is capable of thinking herself as occupying a variety of locations in the allocentric order, can explain how this is possible. But having a conception of space imposes conceptual demands on the perceiver that are unlikely to be required for shape perception. Schellenberg thinks of the required conception of space as "practical". It is not clear, however, what such a practical conception of space amounts to. The next section develops a proposal.

5.2 Public space

The public character of ordinary perceptual objects requires that perceivers transcend their standpoints in individual experience. In the remainder of this section, I explain how this is possible. The core idea is that in ordinary visual space, all locations are presented as standpoints that, were they occupied by co-perceivers, would result in joint attention to the target. A good starting point for this discussion is Kelly's interpretation of Merleau-Ponty's claim that "real" visual objects are seen "from everywhere".

> …the house itself is not the house seen from nowhere, but the house from everywhere. The completed object is translucent, being shot through from all sides by an infinite number of present scrutinies which intersect in its depths leaving nothing hidden" (Merleau-Ponty (1945/2002); quoted after Kelly (2004, p. 91).

As Kelly (2004) explains, the real object is not something a perceiver ever encounters. It is the thing as it would, impossibly, be seen from all perspectives, and the experience of it from any particular viewpoint presents the thing as deviating from how it really is. What enables this indeterminate presence of the real object is the normative character of the background against which it is presented. The background presents the object "in terms of how it ought to change to see the object better" (Kelly, 2004, p. 95). This consideration holds for shape as well as colour: just as there is optimal lighting for seeing an object's colour, so there is an optimal way of seeing an object's shape. We can, on this view, see objects from standpoints we don't occupy because standpoints are not defined by the availability of visual information. Rather, a standpoint reveals what is missing from the presentation of the real object. Kelly interprets this notion of the presentation of an absence by appeal to motor intentionality. He quotes from Merleau-Ponty:

> To look at an object is to inhabit it, and from this habitation to grasp all things in terms of the aspect which they present to it. But insofar as I see those things too, they remain abodes open to my gaze, and, being potentially lodged in them, I already perceive from various angles the central object of my present vision. (Merleau-Ponty (1945/2002); quoted after Kelly (2004, p. 92)).

To "inhabit" an object is, on Kelly's interpretation, for the body to adapt to its shape in the anticipation of action. We experience the environment "in terms of the activities it leads us to perform". Since these activities consist in the manipulation

of whole three-dimensional objects, bodily readiness for action exceeds the agent's visual standpoint, and it is hence that the object's visual aspect that is hidden from view is present in experience, but as an absence. Perceivers transcend their viewpoints on the objects of visual experience because in their skillful coping with the environment they display bodily readiness to take up a different point of view than the one they are currently occupying. In Kelly's example, you may be bodily ready to pick up a mug whose handle is invisible to you by reaching round the mug and gripping the handle from behind. In this anticipation of objectual features that are at present occluded from the agent's view, the visual perceiver is *now* experiencing the object's backside, but as an absence.

A perceiver who experiences objects against a background in Merleau-Ponty's sense treats locations not occupied by herself as origins of motor action and thus as standpoints. These standpoints have to be integrated into the experience of the whole object: they have to be such that, were they occupied by a perceiver, they would enable that perceiver to treat the location of the current perceiver as a standpoint on the object of her own experience. I argued in the previous section that the perceiver has to use relational rather than monadic notions to describe the location of the target relative to standpoints other than her own, since there would otherwise be no way of knowing that it is the same object that is being perceived from these standpoints. The same is true for Merleau-Ponty's account: there has to be an answer to the question of what guarantees that it is the same object that is being singled out from all possible standpoints on the thing. An appealing proposal is that the perceiver experiences visual objects in a spatial format in which all background locations – thus the places not occupied by herself or the target object—are standpoints whence the target's location can be socially triangulated so that, were these standpoints occupied by co-perceivers, joint attention to the target would result. The perceiver thus knows how she would have to move to point out the object to the occupants of these standpoints; and she also knows how she would have to move to point the thing out to the occupant of her own standpoint from standpoints other than their own. She thus knows how she would have to move, from various standpoints, to socially triangulate the location of the target and thereby bring about joint attention to the target, were these standpoints occupied by co-perceivers.

The resulting account of publicity in objectual experience is as follows. For views on which shape experience is explained by appeal to standpoint transcendence in the phenomenal present, shape perceivers are also joint perceivers. Such perceivers operate in a spatial format in which all locations are standpoints, potential origins of action. If a location is a potential origin of action, it is also a potential origin of joint attention: it is an implication of my argument that for theories that appeal to standpoint transcendence to explain shape perception, the two cannot be prised apart. The publicity requirement is then necessarily met for creatures that operate with a spatial framework in which locations other than their own serve as standpoints. It is met because transcending one's standpoint in the phenomenal present requires that one know how to move to point out the object to occupants of standpoints other than one's own and thus achieve joint attention to the target, should these standpoints be occupied by co-perceivers. The notion of public space helps substantiate both Schellenberg's "alter-ego-centric space" and Merleau-Ponty's "view from everywhere".

6

I argued that on the view that shape experience requires standpoint transcendence in the phenomenal present, perceivers who can see shapes are necessarily also capable of joint attention. But, as discussed in Section 8, not all action-based theories of shape perception appeal to standpoint transcendence. My argument does not provide an answer to the question of whether shape perception is necessarily standpoint-transcendent. If you take it, with Grush, that shape perception is possible from within the standpoint of the perceiver, then shapes can be perceived by creatures who don't experience target objects as public; shape perception does not have to be tied to standpoint transcendence. But if you think it isn't, you need an account of how standpoints are transcended in experience when explaining their public character. I have argued that this can be done by appeal to the twin notions of social and public space and the action-based account of joint attention that underwrites them.

Declarations

Conflict of interests There is no conflict of interests that would need to be declared.

References

Avramides, A. (2001). *Other minds*. Routledge.

Beck, S., Riggs, K., & Burns, P. (2011). Multiple developments in counterfactual thinking. In C. Hoerl, T. McCormack, & S. Beck (Eds.), *Understanding counterfactuals, understanding causation: Issues in philosophy and psychology*. Oxford University Press.

Briscoe, R. (2008). Vision, action, and make-perceive. *Mind & Language, 23*(4), 457–497.

Briscoe, R., & Grush, R. (2020). Action-based Theories of Perception. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/sum2020/entries/action-perception. Accessed 8 July 2022.

Campbell, J. (1994). *Past, space, and self*. MIT Press.

Campbell, J. (2002). *Reference and consciousness*. Oxford University Press.

Campbell, J. (2011). An object-dependent perspective on joint attention. In A. Seemann (Ed.), *Joint attention: new developments in psychology, philosophy of mind, and social neuroscience* (pp. 415-430ß). MIT Press.

Carrasco, L. (2011). Visual attention: The past 25 years. *Vision Research, 51*, 1484–1525.

De Vignemont, F. (2018). Peripersonal perception in action. *Synthese, 198*, 4027–4044.

De Vignemont, F. (2021). A minimal sense of here-ness. *The Journal of Philosophy, 118*(4), 169–187.

Eilan, N., Hoerl, C., McCormack, T., & Roessler, J. (Eds.). (2005). *Joint attention: Communication and other minds*. Oxford University Press.

Evans, G. (1982). *The varieties of reference*. Oxford University Press.

Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition, 17*, 535–543.

Gallagher, S., Martinez, S., & Gastelum, M. (2017). Action-space and time: Towards an enactive hermeneutics. In B. Janz (Ed.), *Place, space and hermeneutics* (pp. 83–96). Springer International Publishing.

Grush, R. (2001). Self, world and space: The meaning and mechanisms of ego- and allocentric spatial representation. *Brain and Mind, 1*, 59–92.

Grush, R. (2004). The emulation theory of representation" motor control, imagery, and perception. *Behavioral and Brain Sciences, 27*, 377–442.

Grush, R. (2007). Skill theory v2.0: Dispositions, emulation, and spatial perception. *Synthese, 159*, 389–416.

Hobson, P., & Hobson, J. (2005). What puts the jointness into joint attention? In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 185–204). Oxford University Press.

Kelly, S. D. (2004). Seeing things in Merleau-Ponty. In T. Carman & B. N. Hansen (Eds.), *The Cambridge companion to Merleau-Ponty* (pp. 74–110). Cambridge University Press.

Macpherson, F., & Platchias, D. (2013). *Hallucination: Philosophy and psychology*. MIT Press.

Matthews, W., & Meck, W. (2016). Temporal cognition: Connecting subjective time to perception, attention, and memory. *Psychological Bulletin, 142*(8), 865–907.

Merleau-Ponty, M. (1963). *The structure of behaviour*. Beacon Press.

Merleau-Ponty, M. (1945/2002). *Phenomenology of Perception*. Routledge.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Vaughan Van Hecke, A., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Development, 78*(3), 938–954.

Noe, A. (2004). *Action in perception*. MIT Press.

Noe, A. (2005). Real presence. *Philosophical Topics, 33*(1), 235–264.

Racine, T. (2011). Getting beyond rich and lean views of joint attention. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 21–42). MIT Press.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1997). The space around us. *Science, 277*, 190–191.

Schellenberg, S. (2007). Action and self-location in perception. *Mind, 116*(463), 603–632.

Schwenkler, J. (2014). Vision, self-location, and the phenomenology of the 'Point of View.' *NOUS, 48*(1), 137–155.

Seemann, A. (2011a). Joint attention: Toward a relational account. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 183–202). MIT Press.

Seemann, A. (Ed.). (2011b). *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience*. MIT Press.

Seemann, A. (2019). *The shared world: Perceptual common knowledge, demonstrative communication, and social space*. MIT Press.

Trevarthen, C. (1980). The foundations of intersubjectivity: Development of Interpersonal and Cooperative understanding in infants. In D. Olson (Ed.), *The social foundations of language and thought: Essays in honor of J.S. Bruner* (pp. 316–342). Norton.

Zahavi, D. (2011). The experiential self: Objections and clarifications. In M. Siderits, A. Thompson, & D. Zahavi (Eds.), *Self, no self? Perspectives from analytical, phenomenological, and Indian traditions* (pp. 56–79). Oxford University Press.