JOINT PERCEPTION NEEDS REPRESENTATIONS

Ophelia Deroy (LMU, Faculty of Philosophy & Munich Centre for Neuroscience; University of London Institute of Philosophy, School of Advanced Study) Funding OD is funded by the Volkswagen Foundation (Co-Sense), the Bavarian Research Institute for Digital Transformation (CoLEARN project - Project ID KON-21-0000048).

Louis Longin (LMU, Faculty of Philosophy) LL was funded by the Bavarian Research Institute for Digital Transformation (bidt).

Abstract

Perception often occurs in the presence of others. Here we introduce joint perception as what occurs when the perceptual states of perceivers get mutually entangled, irrespective of the occurrence of joint attention. From a conceptual perspective, this entanglement results in two or more perceivers being aware that objects or scenes are mutually seen and at least visible. From an empirical perspective, this entanglement also results in a difference in how the perceivers see the objects or scenes in question. Here we argue that the empirical evidence of a social sensitivity of perception is best explained by what happens to visual representations and offer joint perception as a place where visual representations play both an explanatory and constitutive role. We still point at difficulties for representationalism to explain the subject of perception when ‘we see’ an object, instead of seeing it alone.

KEYWORDS: Joint attention, joint perception, joint action, visual representation, collective awareness, social influence, cognitive penetration

==========

1. Introduction

Is perception representing the world to us or presenting us with it? The question divides representationalism and non-representationalism about perception. Most of the debates concern what ‘representing’ or ‘presenting’ entails, but our question is different: it asks about

‘us’, and what happens to perception when the plural pronoun is taken literally. When we are perceiving something together with others, rather than alone, do we have a shared or common perceptual representation of the world, and how is this sharing supposed to work? Is it better to think that the world is perceptually presented to several of us?

Asking about perception for several individuals can sound odd at first: if anything, perception is what most theories consider as private, which would suggest it only makes sense at the individual level. Empiricist accounts of sensations, such as Locke’s or Hume’s, and current psychological and philosophical theories may disagree when it comes to the nature of perception or the existence of unconscious perception, but all construe perception as an individual mental state. Sense data are also defined as private, so that they can be related to other sense data for one individual, but not between individuals. Relational accounts of perception more generally tend to define perception as the relation between an isolated perceiver and mind-independent objects in the world.

The privacy claim is both epistemic (nobody can know what your perception is like) and metaphysical (nobody can share your perception). As summarised by Thomas Raleigh “It is commonly accepted that a token experience or sensation always necessarily belongs to some specific subject—and hence that it is metaphysically or logically impossible for another subject to possess one of my experiences or my sensations.” (Raleigh, 2017, p. 639). The claim is not specific to perception, but to all experiences - as pointed for instance by Micheal Tye:

... the pain you are feeling – that particular pain – is private to you. It is yours alone, and necessarily so. No one else could have that particular pain. Of course, conceivably someone else could have a pain that felt just like your pain, but only you could have that very pain. What is true for this one pain is true for pains generally. Indeed, it is true for all mental objects of experience. None of these items of experience can be shared. I cannot have your visual images or feel your tickles, for example. Your images and tickles necessarily *belong to you*. (Tye, 2007, p. 24, emphasis added)

Privacy for the mental objects of experience, here does not mean that *the real object* as perceived cannot be public: by contrast with sensations, where I can clearly not feel the same tickle as you, I can certainly see the same object as you if I am looking at it. Consider an example from Lewis Carroll’s Alice in Wonderland: the mad hatter tea party. After falling down the rabbit hole and entering the garden, Alice meets a mad hatter, and they have a tea party. Alice and the mad hatter are seated at a large table with a teapot situated in the centre of the

table. While their sensations of tasting the tea etc are necessarily private, their targets of attention must not be. If Alice points at the teapot, she expects that
(a) the hatter sees the teapot
(b) he is aware that she, Alice, sees the teapot too (Figure 1).

This type of situation - when two or more persons see the same public object and realise that they are both seeing the same object - is extremely common. Philosophically however, it is a source of problems.

Figure 1: Alice and the hatter jointly attend and perceive the teapot (image created from prompt on D.ALLE)

Problem 1: Do we see things in the same way?

A first problem - in the history of philosophy at least - comes from wondering how similar different people perceive the same object. Alice could see the teapot as being dark green, and the hatter could also agree with that label, but Alice's qualitative experience of dark green may not correspond to what the hatter qualitatively experiences when he looks at the teapot. Perhaps his experience is similar to the experience Alice has when she sees a rose, for instance. Two or more individuals can converge on the same perceptual judgement, yet not have the same perceptual experience. This problem, exemplified by the inverted spectrum thought experiment proposed by Locke, has inspired many subsequent philosophical papers.

Yet the existence of individual differences in how an object is experienced does not threaten the fact that there is an object that we perceive in common: it challenges the fact that one can take one's experience as qualitatively similar to other people's experiences. Nothing in this problem however rests on the fact that the individuals share the same context or look at the object at the same time.

Problem 2: How can we ever look at the same thing?

The second problem discussed by philosophers targets more directly the sharing of the same perceptual context. When Alice sees the teapot and points at it, the hatter can observe the direction of her pointing and gaze, and realise that Alice sees the teapot and wishes him to notice it as well. The situation is different from the situation where Alice is looking at the teapot and realises that the hatter is also looking at it, but also that the hatter himself is not aware that Alice is looking at the teapot. In this case, Alice also realises that the teapot is perceived by her and the hatter, but this is not mutual. In other words, for Alice, it is clear that the teapot is co-perceived, but it is also clear that it is not jointly perceived (see Deroy et al. (2023) for further discussion of the distinction).

How does joint perception work? Following Schiffer et al. (1988), if this is a case of mutual knowledge, Alice will ascribe to the hatter the belief that he sees the teapot, the belief that he believes that she sees the teapot, the belief that he believes that she believes that he sees the teapot, etc. The debate then turns on how one avoids the further iterations that many feel should follow to explain mutual knowledge. The problem of joint attention is to explain how something in the perceptual situation and coordination and observation of attention can put an end to this possible infinite regress. By contrast with the previous problem of individual differences in perception, what matters here is not primarily that people experience the object in similar ways, but how each individual can become aware of the other's private mental state, and of their mental states having the same mutual reference point.

New Problem: Is my perception different because we both perceive together?

The problem we have in mind in this paper is slightly different from those raised by joint attention or inverted spectrum cases, though it could relate to both: what we want to understand is whether Alice's perception of the same teapot is exactly the same if it takes place when she is alone, or when the hatter sits with her and looks at it.

Conceptually, we want to show that what really differs between seeing an object alone or with others is a form of mutual awareness that 'we' see the same object - something which can be related to joint attention, but also can come through other means (section 2). To explain this mutual awareness, both representational and non-representational views of perception can perform as well (section 3). The challenge however comes from empirical evidence which suggests that what is perceived is not exactly the same when the same perceiver sees the same object either alone or with others: When Alice is with the hatter, she may see the teapot differently (section 4). This difference, we argue, can best and even only be accommodated by positing visual representations (section 5).

2. Joint perception from the armchair: Joint visibility

Attention, but also action, are meant to give rise to new kinds of mental states when they involve coordinated agents: joint attention and joint action are two phenomena which suggest that our minds sometimes meet. For representationalism, this means that we somewhat share the same representation and know that we do; for non-representationalism, this must mean that we partake in the same attentional episode or action, and relate to the same object.

Our target here is the phenomenon of joint perception, which relates but is also broader than joint attention and joint action. To insist first on how they relate, it is clear that joint perception is the necessary result of joint attention: once several agents have jointly attended to a common object, they also both see the target together, and are aware that they both see it. When Alice and the hatter jointly attend to the teapot, they are basing their interaction and discussion ("what a lovely green teapot") on the assumption that they can both see the teapot in question, including its colour, its shape, etc. Joint attention goes hand in hand with joint perception: When joint attention occurs, it is both practically and rationally expected that joint perception also occurs. On the other hand, when several agents coordinate their action on a common, concrete object, they need also to perceive it in common. This is true in practice, and this is necessary for rationally justifying the actions of the individuals. If Alice and the hatter agree to jointly lift the teapot, because it is heavy and fragile, then the two characters base their action on the assumption that they can both see the shape of the teapot, as well as feel its weight, etc.

In both cases, the teapot is common not only as an object of shared attention or the target of a coordinated action, but also as a shared perceptual object: each can mutually and rationally expect the other to see what they see.

Joint perception is then not just the co-occurrence of individual perceptions, which could look like this in the case of the teapot:

Co-occurrence of perception
(Alice's individual perception) Anna sees the teapot.
(Hatter's individual perception) The hatter sees the teapot.

To count as a case of joint perception, the case must at least follow roughly the following lines:

Joint perception
Alice and the hatter are mutually aware that they are seeing the same teapot.

This sounds really close to joint attention: When two or more people overtly focus on the same object at the same time, with each being aware of the other's interest, something specific happens which is more than the juxtaposition of their individual attention (independent occurrence) but also more than one locus of attention taking the other attention into account. What is making this a case of jointness is the mutual realisation by the characters that they are both attending to the same thing, which can be captured roughly along these lines:

Joint attention
Alice and the hatter are mutually aware that they are both attending to the teapot.

The point is not here to exactly see how this mutual clause should be expressed - which is a topic for the field of joint attention (see Campbell (2018); Battich and Geurts (2020) for discussion), but understand what the jointness of attention presupposes. For joint attention, the target of attention is not just objectively the same, but also part of what is usually known as "a common ground": Alice and the hatter can mutually and rationally expect the other to know what the other focuses on, and sees and refers to when they say, "the teapot", or simply use the pronoun "it".

Joint perception however is more extensive: Alice could be aware that the hatter is seeing the teapot, while also being aware that he is actually looking at the cup near to the teapot. In this case, there is no shared object of attention, but there is a shared object of perception. In the same respect, the hatter could be aware that Alice is seeing the teapot, though he is staring more precisely at the cup. Joint perception here can occur without the mutual entanglement of joint attention.

It is also possible that joint perception occurs not because of joint attention, but because the hatter is behind Alice and tells her that he sees the teapot that is in front of her. Here, Alice is aware that the hatter sees the teapot, and knows that she sees the teapot - and there is mutual knowledge in this case, but it does not occur because of a coordination of attention It could, in this respect, account for what happens online when two people read a shared document at the same time, and see each other online: no coordination of gaze is involved, but both are aware that they see the same document.

Expanding the case mentioned earlier, it could be also the case that Alice notices that the hatter looks at the teapot, and realises that they are seeing the same teapot, but Alice does not expect the hatter to realise that they look at the same teapot. In this case however, the clause of joint perception fails, and reduces to co-occurrence of perception plus some awareness of the object of the other's seeing. So, while joint perception is not tied to joint attention to the same object, it is different from the mere realisation that the object is visible to someone else that oneself - something which is a more general co-perception (Deroy et al., 2023). What matters is that the joint perception is different from the public character of the visible object, in that it is tied to certain viewers' entanglement. In other words, there is more to it than the awareness that one's object of perception is independent of one's mind, and that it can be perceived by others.

We spoke so far of joint perception as a possible carry-over of joint attention, or the result of other contextual coordination. But does it need to belong to perception? In other words, could this mutual awareness of seeing the same object not be a post-perceptual stage, where perceivers are inferring from gaze or other cues that they must be seeing the same object. We need to agree that there is a difference between 'joint perception' and 'post-perceptual judgement that perception is shared'. To go back to the example of Alice, there is a difference between joint perception and a perceptual judgement of sharing:

Joint perception
Alice and the hatter are mutually aware that they are seeing the same teapot.

Judgement that visual perception is shared
Alice and the hatter are mutually judging that they are seeing the same teapot.

Nobody is denying that a judgement that perception is shared can also occur when two agents, capable of judgement, are seeing the same object, and are aware that they do. The judgement

is, however, not necessary. Empirical studies (see section 4) have shown that the presence of another person viewing the same object can permeate perceptual processing prior to forming a perceptual judgement. In other words, shared perception not only affects *what* participants report but also *how* they perceive their environment. For example, in some recent studies (Germar et al., 2019, 2023), participants were presented with stimuli panels filled with orange and blue dots. They were asked to indicate whether the target stimulus was dominated by the colour orange or blue and adjust the proportion of orange and blue dots until they perceived their proportion as being equal. Participants then received social feedback on what other participants had done. Comparing the effect of social feedback on perceptual processing with EEG and perceptual reports reveals that only the perceptual processing was affected by social information about others' perceptual decisions. Germar et al. findings bolster the point that viewing an object in common does not influence perceptual judgements but can also affect perception itself (see also Molleman et al., 2019).

We can conclude that the state corresponding to a situation where Alice sees the teapot alone is different from the state where she and the hatter jointly see the teapot. The next question is whether this state - where the teapot is part of joint perception - is best captured by a representationalist framework.

3. Joint perception from the armchair: the role of representations

3.1. Can representations explain joint perception?

Many researchers have argued, or accepted, that joint attention requires the use of representations, and may even require the use of meta-representations of the other's mental state. In joint attention, each character represents the fact that the other is also attending to the same object and that they know that they are both attending to the same object. The object is then used to update cognitive representations (beliefs) that are considered common knowledge by each individual.

Shifting the debate to joint perception - understood as a state where the perceivers are aware of the joint visibility of the object - highlights however another issue: In joint attention, what tends to be stressed is the commonality of knowledge or belief, where agents are aware that an object is part of a 'common ground' ; with joint perception, what comes in the foreground is the commonality of a certain perceptual object in context.

Accommodating this sharing and the joint visibility clause in a representationalist framework means, at least as a first approximation, proposing something along the following lines to explain the case of Alice when she jointly perceives the teapot with the hatter:

Joint perception (representationalist version)
Alice is representing that "we" (she and the hatter) are visually representing the same teapot.

The first explanatory role of representation relates here to classic 'mind-reading' capacities, more particularly the capacity to represent what other agents represent. Whereas mind-reading traditionally focuses on what other agents represent at the level of beliefs, what is at stake here is the visual representation of the scene or object. A second role of representations here would be to explain then how the perceivers shift from a visual representation of the object as visible to only to them, to a visual representation where the object is visible to them and the other(s): from a first-person point of view, Alice would no longer see the teapot as 'visible to me' but 'visible to us' (a "collective", see Shteynberg et al., 2023).

This issue taps into two related debates: first, whether and how the subject of perception - individual or collective - appears in perceptual representation, and second, whether and how it makes sense to posit a collective agent of perception.

Not all proponents of representationalism would commit to say that, when Alice sees the teapot alone, she is visually representing the teapot, and also representing the fact that she is the one seeing that teapot. What one thinks about the subject of representation figuring in the representation will have consequences on how to handle the upgrade from one to more than one subject. Take for instance a self-representational theory of perceptual consciousness, of the kind proposed by Kriegel (2009), which proposes that when I have a visual experience, I visually represent an object, and that this visual experience also represents itself as 'for-me'. The subject is here part of the content of the representation - at least in a phenomenological sense.

Can a self-representational account eventually be adjusted to accommodate a plurality of subjects, with the visual experience representing itself as for-us, rather than for-me? The problem here is that the exact content of individual visual experience may not be shared: Alice may represent the teapot from one spatial perspective, and the hatter from another, and it becomes difficult to say which visual representation of the tea is both for-Alice-and-for-the-hatter. The move from single to collective in non-perspectival contents is not problematic, as it is the same content that one can attribute to themselves individually and collectively. The move

from single to collective in visual perspectival contents, is more problematic. At least Alice would be fallible in representing the teapot as she visually represents it from her perspective as being also jointly visible to her and the hatter. What she could accurately represent as jointly visible is no longer the visible object but a reconstructed object, which somehow integrates her and the hatter's perspective.

How about theories which consider that the subject of representation appears not in the content but in the mode of the representation? Some theories indeed appeal to a Fregean mode of presentation, which avoids having to push the subject in what is represented. For joint perception, the idea would be that the mode of presentation for a given content would be singular when Alice sees the teapot by herself, and plural when she sees it jointly with the hatter. (me-mode versus we-mode, see for instance Gallotti & Frith, 2013).

A possible issue here will be self-representational accounts and dominant interpretations of the we-mode accounts suppose a phenomenal difference between perceptually representing the same object or scene alone or with others, which needs to be fleshed out. For people like Kriegel, it is not the ontological subject that is self-represented in perception, but a phenomenal subject 'I' (the subject as given in the mode of the experience). The phenomenological subject need not always coincide with the ontological subject: the phenomenological subject can be a "we", a plural subject rather than an "I". Yet positing a phenomenological sense of we-ness in joint perception is controversial (Zahavi, 2018), if it is solely to accommodate common sense statements like 'we see the teapot' and cannot really connect to a felt experience or empirical evidence. As such, it can become a costly theoretical commitment.

Alternatively, joint perception can be accommodated by other representationalists accounts, which do not demand that the subject of perception is represented in the content or present in the mode of presentation of the representation. We can think of several. Using a mental files framework, one could say that the perceptual representations formed during joint perception get tagged to a 'we' mental file instead of a 'I' mental file. Using functionalist theories of representations, the representations formed during joint perception are used later differently by other cognitive systems. In these views, the perceptual awareness that Alice is not alone seeing the teapot or the background belief that she is seeing it with the hatter can be sufficient to explain why the perceptual representation of the teapot will have a different 'fate' following joint perception than it would have if Alice was aware of being alone. For instance, after joint perception, she could refer to the teapot when speaking to the hatter and ask, "Do you still have that lovely green teapot?". Or if she breaks the teapot while carrying it alone to the kitchen, she is aware that the hatter will later remember having seen this very teapot and notice

it is missing. She could, more immediately, realise that she needs to be fast if she wants to grab the teapot before the hatter does.

Here the difference between single and joint perception comes from the function or type of input-output that the visual representation has for the perceiver. One problem here however comes from the fact that such cognitive differences in function are not necessarily accessible (or at least there is no reason to think they are) to the subjects of perception themselves: there is no perceptual difference for Alice between the case where she sees the teapot alone or with the hatter.

This is where the account needs to insist on a key difference between the fact that a common ground is *manifested* during judgement and the fact that it is *established* by judgement. Compare the two following scenarios:

1. Alice meets the hatter, and she tells him “I like the green colour of the teapot that they have in this shop”. Later, the hatter goes to the shop, sees the green teapot and reports to Alice “I like that green too”.

2. Alice sits in front of the hatter and they both look at the teapot with interest. They both are aware of the other and realise that they both see the same teapot. At some point the hatter says, “I like that green too”.

In the first case, judgement and conversation establish the common ground, which is then manifested in the hatter’s utterance. In the second, perception establishes the common ground, and the utterance of the hatter’s manifest it.

Representationalist frameworks, which can accommodate the common ground established verbally, are also well placed to capture this commonality established in perception. But what about non-representationalist accounts?

3.2. Joint visibility without representations?

Some philosophers, notably John Campbell, resist the representational interpretation of joint attention, arguing that representations are not necessary to make sense of the mutual awareness of the attending agents. What happens, in this non-representational version, is that the commonality of the attended target is “in the open”: as Alice and the hatter jointly attend to the object, staring and pointing at it, it is directly clear to them that the object is attended by the

two of them - they don't need to represent that commonality at any level. According to Campbell, one can dispense with representations and still make sense of the jointness of attention.

Relatedly, like for joint attention, some people have argued that joint action can, at least in some early or basic forms, proceed without representations. For instance, Talamanca (2021) argues that joint action can occur in some non-human primates as a form of socially embedded skilful unreflective action, as understood by Rietveld (2008): chimpanzees in the Taï forest in Côte d'Ivoire can jointly coordinate their actions during hunting, without having to represent others' goals. Relatedly, Satne has argued that joint action, like joint attention, can be successfully explained through external facts such as roles and affordances: her view, she writes, is not just "that minimal cooperation does not require a representational state about being cooperative. It is the stronger general claim that collective intentions, recognition, acceptance, and joint commitments are not necessary for minimal cooperation at least when individuals are members of an organised group" (Satne & Salice, 2020).

In summary, joint attention and joint action, though well explained by positing representations, are currently being revisited in a non-representational framework. What not then joint perception? After all, translating the concept of 'openness' from joint attention to joint perception could take care of the joint visibility, and coordination of perceptual roles or affordances could account for joint perception, especially if it supports coordinated action in an unreflective manner.

Non-representational accounts of joint perception here would seem to face much less difficulties upgrading from the individual to the collective than representational accounts, which struggle with the joint visibility clause. The situation changes however, once we turn to other aspects of joint perception, which go beyond joint visibility and offer a more complex picture of joint perception.

4. Joint perception: Empirical evidence

We want to emphasise the importance of incorporating third-person empirical evidence to understand what happens when perceiving something alone versus with others. This distinction is challenging to grasp subjectively for a couple of reasons. Firstly, the differences between these two perceptual states can be subtle and easily overlooked. Secondly, experiencing something alone or in a group is confined to one perspective at a time. Thus, an individual cannot simultaneously occupy both scenarios to make a direct or fair comparison.

Below we turn and discuss different kinds of empirical evidence which are better equipped at performing such comparisons, and the surprising results they lead to: that perception operates differently when the same perceiver sees the same object either alone or together with one or more perceivers. Before we do so, it is important here to remember two core differences between looking at vision from a philosophical and from an experimental perspective. In experiments, what is measured is not perceptual experience, but *perceptual decisions.* Decisions are mostly recorded through people pressing on a button to indicate which of two alternatives correspond to what they perceptually grasped. One needs to turn to other evidence to better infer whether this means that they were having a conscious perceptual experience, and of what kind.

Additionally, the stimuli are rarely framed to measure the things common sense attributes to perception, like seeing a green teapot on a table next to a cup, but *one specific aspect of perception*: how well people can visually detect whether something is present or absent, find an object in a scene, discriminate between two different contrasts, or categorise and identify something as being of a given kind. Keeping these differences in mind, we can look at what empirical studies show regarding how specific perceptual decisions occur when people are jointly perceiving an object.

4.1. Perceptual processing is faster during joint perception

The first type of evidence pertains to the speed of decision-making in cases of joint perception. It has been observed that people can detect and recognize objects more quickly when these objects appear in the location where another person is looking. This increase in speed is contingent upon the ability to observe where the other person is gazing (Friesen & Kingstone, 1998; Driver et al., 1999). Others have extended the evidence for the perceptual cueing of gaze direction to human body postures (Azarian et al., 2017). Compared to pure gaze direction experiments, the body posture has a similar effect on perceptual decisions compared to gaze cueing: when the body posture is congruent with the identifiable target, participants were faster to detect the target than when the body posture of the presented agent is incongruent with the identifiable target. Note here that most of these experiments rely on instructing the participant to look in a certain direction on a screen, for instance the left side, and presenting an avatar on the same screen. The avatar could be looking in the same direction as the one the participant, or in another direction, creating either a situation of joint perception or not. It's important to note that joint perception in these experiments doesn't necessarily equate to joint

attention. Participants aren't informed that the avatar is focusing on the same targets as they are. The effect is based solely on the avatar's body and eye orientation.

The fact that people are faster at detecting and recognising visible targets when someone else is seeing them is shown to be highly automatic, akin to a reflex: it is fast and occurs without control. Observers continue to follow gaze even when the gaze cue is entirely non-predictive of where the target will appear, and thus is detrimental to performance (Friesen & Kingstone, 1998). This is at least some evidence that what happens is somewhat perceptual, rather than cognitive - at least if cognitive means reflective and partly under control.

Some debates and studies problematically seem to show however that this speeding up is not specific to joint perception: symbolic cues such as arrows and directional words also reliably orient attention across a similar time course (Taylor & Klein, 2000; Hommel et al., 2001; Ristic et al., 2002; Tipples, 2002, 2008). However, the mechanisms underpinning the effects of gaze and arrows have also shown to proceed differently: eye gaze cuing triggers focussed activation related to enhanced visual processing, while arrows activate a much broader network including areas specifically related to volitional orienting (Hietanen et al., 2006).

Granting then that there is something special in the case of joint perception, another problem arises. What do changes in speed tell us psychologically, or philosophically about perception? What do a few hundred milliseconds really change for the agent? Is this a direct perceptual effect, or rather an indirect effect whereby more aroused and motivated agents press the response button faster? In other words, is there a difference in how the target is perceived, or just how fast people respond to the perceived target?

A long tradition of experiments warns against a mere social facilitation of responses: when participants are asked to perform a task with someone else present in the room, even more if that someone is known to perform or having performed the same task, they easily feel under pressure to compete, meaning they respond faster (and often less well). The mere physical presence of someone in the room can also change their physiological levels of arousal, and make them respond faster. Two types of arguments suggest that at least *some* of the speed difference corresponds to a difference in perceptual processing, rather than a tendency to press response buttons faster because someone is around: first, as said above, the speed difference corresponds to enhanced visual processing during joint perception, which can be tested with neuro-imagery ; second, it is sensitive to the gaze cues, and not to what someone else is doing: If the effect was all about responding as fast, for instance because of social comparison and a motivation to be quicker than, or as fast as, the other, then just seeing that

someone moves to press the button should do it. Yet Friesen et al. (2004) have ruled out that this is the case, and demonstrated that the reflexive orientation to other's gaze direction is attributable to the observed gaze cue and not the mere onset of another agent.

Another argument, evidenced in Battich et al. (2021), is that faster responses in joint perception are not less accurate than slower responses in individual perception. In this study, people were presented with brief flashes of light, and asked to say how many were flashed. Following the famous 'flash illusions' (Shams et al., 2000), the flashes were also accompanied by sounds: hearing one sound could sometimes make people see one single flash, when two were presented (a 'fusion' illusion) while hearing two sounds could sometimes make people see two flashes, when actually only one was presented ('a fission illusion'). People would then have a certain error rate in their visual perceptions, depending on how often they counted 1 flash instead of 2, or 2 instead of 1. Crucially for our current argument, the same people were asked to perform the same task alone and jointly with someone else: both would look at the same screen, hear the same sounds, and be asked to say how many flashes they saw. People were faster at responding when perceiving together with someone else, and their error rate was comparable to the one they had when perceiving alone. If one was simply responding faster because of social comparison, we would expect that their responses would also be less accurate, because of a speed-accuracy trade off. The fact that such a trade-off does not occur is a strong argument in favour of joint perception being different (and in this case, more efficient) than individual perception.

4.2. Visual perspective is different during joint perception

Other facets of visual perception are also influenced by the presence of another person, specifically by the perspective they hold on to a given scene. A pivotal study by Samson et al. (2010) demonstrates that participants struggle to disregard another person's viewpoint when both individuals are observing the same object. This phenomenon occurs not only when the other person's spatial perspective is relevant to the task being performed but also when it has no bearing on the task at all, as shown in research by Böckler and Zwickel (2013). This indicates that the process of understanding what someone else sees happens involuntarily and is largely automatic. Such findings suggest that this interference in perception is less about conscious judgement and more about an automatic integration of another person's perspective into one's own perceptual processing.

Freundlieb et al. (2017, 2018) delved into the spontaneous occurrence of such perspectival integration when the other's perspective is significant for mental tasks rather than possible physical actions. In their experiment, participants were seated at a 90° angle relative to a confederate and engaged in a task involving the semantic categorization of written words. For the participants, the words were always displayed vertically, but for the confederate, their orientation varied—appearing either right-side-up or upside-down—depending on the confederate's position. Participants were slower in categorising words that appeared upside-down from the confederate's viewpoint, though the words were always "seen" from the same perspective for them. This effect vanished when the confederate's visual access was blocked with opaque goggles.

4.3. Perceptual categorisation is different during joint perception

Is there more evidence that the fact that two or more people jointly perceive the same object means that they see the world differently? This question has been at the forefront of many experiments and discussions, at least since the famous experiment conducted by Solomon Asch in the 1930's. In this famous experiment, many participants reported seeing a line as shorter as they would report, if alone, when placed in a group of people stating that it was indeed shorter. There are good reasons to believe here that the report shows an adjustment at the level of public expression, and not even at the level of private judgement - let alone then at the level of perception. In brief, there is social conformity or social influence in perceptual judgements, especially when they need to be expressed in public. The matter is actually more complicated, with additional evidence now showing that subjective certainty in what we see is different in a private or public setting (Bang et al., 2020).

Still, saying that differences occur at the level of reports, judgements or even subjective certainty does not rule out that differences can also occur in perception. This was recently tested by Zanesco et al. (2019). Their experiment builds on an experiment constructed after Ash's conformity experiments and performed by Moscovici and Zavalloni (1969). There, instead of lines, people would be seeing patches of colour, either clearly blue, clearly green, or in between, and asked to say which colour they saw - green or blue? The same patches of colour were then presented a second time, but this time along with information about the colour that other perceivers had seen. After receiving the social feedback, the participants were asked to say which colour it looked like.

Zanesco et al. (2019) found that social feedback expectedly influences perceptual categorisation when ambiguous colours are presented, but also when distinct colours were

presented. Most importantly, electrophysiological results show that the social feedback influenced early perceptual brain processes and were not only a matter of later reporting. In other words, their results give us reasons to think that Alice sees the same blue-green teapot differently when she sees it with the hatter, knowing at least that he sees it as greener than she does on her own.

Of course, scenarios of this type require that one knows what colour other people perceive. In the experiment conducted by Zanesco and colleagues, this information is not given perceptually: It is provided symbolically, and even with delay. People are told about the perceptual judgement that others formed when they were exposed alone to the same patch of colour.

To make these results relevant to joint perception, the set-up should be substantially adjusted but it seems highly plausible to do so. After all, it is possible that judgements are shared during a discussion while two or more persons are looking at the coloured object, so the delay is not necessary, and the object is also jointly visible. It is also possible for the social feedback to be already known when the two people silently watch the same-coloured object - imagine for instance, that Jim and Jules have disagreed in the past on the colour of a given logo - Jim seeing it as blue, Jules as green. One day, they happen to see the same logo on a billboard as they walk silently and Jules knows that Jim considers it as blue, while alone, he considers it as green. In this case, it is possible to think that Jules' categorisation of the colour would be different. It could be also possible to perceive that someone disagrees with us, by frowning for instance, instead of saying that they have reached a different judgement.

What matters here is the importance of empirically plausible cases where visual categorisation - how a certain colour looks perceptually - can differ when a viewer is alone or with others who have another categorical representation of it.

4.4. Perceptual detection is different during joint perception

Going back to situations where two people are jointly seeing a target at the same time, we can also find new evidence that what is seen is different in joint and single perception. Seow and Fleming (2019) recently showed that we are better at detecting the presence of a faint object when accompanied by another agent. The task consisted in presenting a Gabor patch (or no Gabor patch) close to detection threshold, either on the right or left side of a room that the participant was looking at. In some trials, the target was present, and in some others, the target was absent: The role of the participant was to say where the target was, if they saw something

at all, or else say that no target was present. Crucially, in some conditions, a human avatar placed on the screen was also looking - either to the left, or the right side of the room. In the case where a target would be presented, say on the left, the avatar's gaze direction could be congruent, if it was looking on the left, or incongruent, if it was looking on the right. Participants were better at detecting the target - that is reporting where something appeared, when the target was presented on the side that the avatar was also looking at. This means that the same faint evidence will lead to the representation "something appeared on the left" when someone else is also looking at it, and "nothing was either on the left or the right" when no one is there. Importantly, this effect is not simply due to the facilitation introduced by someone's head being turned to one side - as a control condition is run where a second perceiver is also present, but his eyes are masked - and no such facilitation is then observed. This type of evidence introduces a very important argument for the difference between joint and single perception: an object can be seen in one situation and not in the other.

4.5. Higher-level properties also are different during joint perception

The studies reported above all look at low-level properties which are uncontroversially perceptual: objects and events in the experiments come with differences in location, orientation, shape, colour etc. Not everyone will recognise other higher-level properties, such as for instance aesthetic or evaluative properties (something being harmonious, balanced, appealing, disgusting, etc.) or action properties (something being graspable, liftable, climbable, etc.) as perceptual, but others will. Though a controversial question, we can also see studies which speak to what happens to higher-level properties in joint perception.

Seeing an object together with someone is sufficient to make it more likeable (Bayliss et al., 2006), an effect which is modulated by the emotional expression of the observed face (Bayliss et al., 2007) or the fluency with which the object is reached (Hayes et al., 2008). What is more, the action properties of an object might be modified by the gaze of another person, as shown by kinematic studies using motor interference (Castiello, 2003). Motor interference occurs when a target object is presented along with distractors, for instance when an agent needs to grasp a large ball among smaller balls. Studies show that, if someone sees either the full body, or even just the eyes of an agent performing a different task, then this observer will in turn show a motor interference in grasping a ball, even when the distractors are not present. In other words, just perceiving an object jointly with someone who has a different motor goal in mind can be sufficient to influence one's perception of affordances.

4.6. Interim summary

So far, we have identified joint perception as a kind of mental state that is not reducible to the mere co-occurrence between individual acts or states of perception. More precisely, we can propose that joint perception is characterised not only by joint visibility but also by a form of social sensitivity of perception. We have shown experimental evidence establishing the social sensitivity of perception in five main ways: faster perceptual processing (4.1.), automatic visual perspective taking (4.2.), different perceptual categorisation (4.3.), heightened perceptual detection (4.4.), and altered higher-level perceptual properties (4.5.). Accepting that, in the case of Alice, the following two proposition hold:

Joint perception
Alice and the hatter are mutually aware that they are seeing the same teapot.

Social sensitivity of perception
Alice sees the same teapot differently when she sees it jointly with the hatter, rather than alone.

We can now turn to our main issue: what does this mean for the role of visual representations?

5. Does social sensitivity require representations?

The way we have formulated social sensitivity of perception is not by definition tied to a representationalist or non-representationalist framework. Yet the capacity for the two types of accounts to accommodate this characteristic differs.

People who favour representationalism can easily construct it as a form of 'social penetrability' of perception, by reference to the well-known construct of cognitive penetrability (Pylyshyn, 1999; Deroy, 2013; Siegel, 2017): in the case of cognitive penetrability, the representation that a perceiver forms of the same object when they have a certain belief A about the object is different from the representation they would form if they did not have the belief A (and instead would have no belief or another belief); in the case of social penetrability, the representation that a perceiver forms of the same object when they are with another perceiver X is different from the representation they would have if they were alone. In addition, the framework of social penetrability could explain why the representation that the perceiver forms of the same object when they are with another perceiver X is also different from the representation they would have if they were with another perceiver Y, or more than one perceiver.

The analogy between social and cognitive penetrability however should not suggest that the mechanisms are similar: what is similar is the fact that perceptual representation is causally influenced by another representation, but cognitive penetration considers that this influence needs to be top-down and from a cognitive representation (however one decides to fix the criteria, see Silins, 2016; Deroy, 2013, 2019). By contrast, social penetrability can occur through top-down as well as bottom-up processes, and may not be restricted to cognitive representations in the role of influencers. As said before, some influences may be indeed very automatic, and correspond to a 'natural' adjustment of spatial perspective evolved to assist with coordinated actions (Kampis and Southgate, 2020).

If representationalism can explain social sensitivity of perception, what about non representationalist accounts? Recent attempts to rephrase cognitive penetration as a form of 'social permeability' of perception (following Ramstead et al., 2020) may seem relevant, but they fail so far to provide a clear explanation of how the mechanism works across the board. In the case of perception of action opportunities, for instance, it is possible to think that one's perception of the object - characterised by the relation between the perceived object and the perceiver - is causally influenced by the relation between a co-actor and the perceiver, and the relation between the perceived object and the co-actor. Social perception would then be at least a three-part relation between the perceiver, a (commonly) perceived object, and the co-perceiver: what happens in the case of joint perception would be that the visible presence of someone else, who is looking at the object and therefore can be interpreted as another agent-object relation, affects the other perceiver's relation to the same visible object. So far however, without appeal to representations, social permeability is restricted to observable relations - like spatial relations - and it is difficult to see how it could explain that the colour of an object, or the very presence of an object differs when a perceiver relates to an object alone or with others. Even extending the perceptual relations to involve other features of the perceptual environment cannot capture the full extent of social influence on perceptual processing and perceptual content demonstrated empirically. An increase in perceptual sensitivity (Seow and Flemming, 2019) and a change in neural processing of visual stimuli after social feedback (Germar et al., 2019, 2023; Zanesco et al., 2019) point at a perceptual social sensitivity that goes beyond mapping a relation between a co-perceiver and a commonly perceived object. Instead, in cases of low-level social perceptual influence, the co-perceiver's perceptual content of the commonly perceived object is represented by the perceiver.

Standing on the representationalist camp however, we cannot speak for other resources of the non-representational accounts that could accommodate these situations. Our point is however to offer it as a challenge for those who wish to dispense with representations in perception. It

is also to stress that accounts which accept visual representations do not face such a challenge, and can provide explanations for the social sensitivity of perception reviewed above.

6. Conclusion - Representing the world jointly

Phenomena like joint attention and joint action have already shown that our mental states can be both private and entangled with other people's mental states. The same, we argue, is true of perceptual states.

How to capture the mutual entanglement in joint attention and action precisely is debated, but our focus has been on joint perception. We show that, when the perceptual states of two perceivers (or more) not-only accidentally co-occur but coordinate, for instance to solve a common problem, each individual perceptual state, as it occurs for each participant, also includes and is responsive to this mutual entanglement with others' perceptual states. Joint perception is challenging to accommodate, both for representationalists and non-representationalists, whose accounts are equally focused on individual perception.

Our main point here has been to show that representations do well at capturing and explaining the fact that how one sees an object or scene differs depending on whether they are alone or with others and suggest that representations may be even necessary to account for such a difference. At the same time, representational accounts are not well equipped to think about the subject of representations, especially perceptual ones, being held in common with others.

This conclusion may seem aporetic, and we do not mean to suggest that solutions cannot emerge in the future. Our own representational account, here, is quite radical in that we suggest that joint perception may simply bypass awareness, so that joint visibility is no longer a characteristic of joint perception, as it seems from the armchair. We only leave it as a fact of judgement, or at best, part of a post-perceptual processing. For now, it is sufficient for us to conclude on the fact that joint perception offers a place where representations continue - and should continue - to play a role, and not be easily dispensed with.

REFERENCES

Azarian, B., Buzzell, G. A., Esser, E. G., Dornstauder, A., & Peterson, M. S. (2017). Averted body postures facilitate orienting of the eyes. Acta psychologica, 175, 28-32.

Bang, D., Ershadmanesh, S., Nili, H., & Fleming, S. M. (2020). Private–public mappings in human prefrontal cortex. Elife, 9, e56477.

Battich, L., Fairhurst, M., & Deroy, O. (2020). Coordinating attention requires coordinated senses. Psychonomic bulletin & review, 1-13.

Battich, L., & Geurts, B. (2020). Joint attention and perceptual experience. Synthese, 1-14.

Bayliss, A. P., Paul, M. A., Cannon, P. R., & Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. Psychonomic bulletin & review, 13(6), 1061-1066.

Bayliss, A. P., Frischen, A., Fenske, M. J., & Tipper, S. P. (2007). Affective evaluations of objects are influenced by observed gaze direction and emotional expression. Cognition, 104(3), 644-653.

Böckler, A., & Zwickel, J. (2013). Influences of spontaneous perspective taking on spatial and identity processing of faces. Social cognitive and affective neuroscience, 8(7), 735-740.

Campbell, J. (2018). Joint attention. In M. Jankovic & K. Ludwig (Eds.), *The Routledge handbook of collective intentionality* (pp. 115–129). New York, NY: Routledge.

Castiello, U. (2003). Understanding other people's actions: intention and attention. Journal of Experimental Psychology: Human Perception and Performance, 29(2), 416.

Deroy, O. (2013). Object-sensitivity versus cognitive penetrability of perception. Philosophical studies, 162(1), 87-107.

Deroy, O. (2019). Predictions do not entail cognitive penetration:“Racial” Biases in predictive models of perception. In Limbeck-Lilienau. C. and Stadler. F. (eds) The Philosophy of Perception (pp. 235-248). Berlin: De Gruyter.

Deroy, O., Longin, L., & Bahrami, B. (2023). *Co-Perceiving: Bringing the social into perception*. PsychArchives. https://doi.org/10.23668/psycharchives.13456

Driver IV, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. Visual cognition, 6(5), 509-540.

Freundlieb, M., Sebanz, N., & Kovács, Á. M. (2017). Out of your sight, out of my mind: Knowledge about another person’s visual access modulates spontaneous visuospatial perspective-taking. Journal of experimental psychology: human perception and performance, 43(6), 1065.

Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2018). Reading your mind while you are reading—Evidence for spontaneous visuospatial perspective taking during a semantic categorization task. Psychological Science, 29(4), 614-622.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. Psychonomic bulletin & review, 5(3), 490-495.

Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. Journal of Experimental Psychology: Human Perception and Performance, 30(2), 319.

Gallotti, M., & Frith, C. D. (2013). Social cognition in the we-mode. Trends in cognitive sciences, 17(4), 160-165.

Germar, M., & Mojzisch, A. (2019). Learning of social norms can lead to a persistent perceptual bias: A diffusion model approach. Journal of Experimental Social Psychology, 84, 103801

Germar, M., Duderstadt, V. H., & Mojzisch, A. (2023). Social norms shape visual appearance: Taking a closer look at the link between social norm learning and perceptual decision-making. Cognition, 241, 105611.

Hayes, A. E., Paul, M. A., Beuger, B., & Tipper, S. P. (2008). Self produced and observed actions influence emotion: the roles of action fluency and eye gaze. Psychological research, 72(4), 461-472.

Hietanen, J. K., Nummenmaa, L., Nyman, M. J., Parkkola, R., & Hämäläinen, H. (2006). Automatic attention orienting by social and symbolic cues activates different neural networks: An fMRI study. Neuroimage, 33(1), 406-413.

Hommel, B., Pratt, J., Colzato, L., & Godijn, R. (2001). Symbolic control of visual attention. Psychological science, 12(5), 360-365.

Kampis, D., & Southgate, V. (2020). Altercentric cognition: how others influence our cognitive processing. Trends in Cognitive Sciences, 24(11), 945-959.

Kriegel, U. (2009). Subjective consciousness: A self-representational theory. Oxford University Press.

Moscovici, S., & Zavalloni, M. (1969). The group as a polarizer of attitudes. Journal of personality and social psychology, 12(2), 125.

Molleman, L., Kurvers, R. H., & van den Bos, W. (2019). Unleashing the BEAST: A brief measure of human social information use. Evolution and Human Behavior, 40(5), 492-499.

Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. Behavioral and brain sciences, 22(3), 341-365.

Raleigh, T. (2017). Phenomenal privacy, similarity and communicability. Ergo: An Open Access Journal of Philosophy, 4(22), 637-667.

Ramstead, M. J., Friston, K. J., & Hipólito, I. (2020). Is the free-energy principle a formal theory of semantics? From variational density dynamics to neural and phenotypic representations. Entropy, 22(8), 889.

Rietveld, E. (2008). Situated normativity: The normative aspect of embodied cognition in unreflective action. Mind, 117(468), 973-1001.

Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. Psychonomic bulletin & review, 9(3), 507-513.

Samson, D. et al. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. Journal of Experimental Psychology: Human Perception and Performance 36, 1255–126.

Satne, G., & Salice, A. (2020). Shared intentionality and the cooperative evolutionary hypothesis. In Minimal cooperation and shared agency (pp. 71-92). Springer, Cham.

Schiffer, S. (1988). Review of The Varieties of Reference [Review of Review of The Varieties of Reference, by G. Evans & J. McDowell]. The Journal of Philosophy, 85(1), 33–42. https://doi.org/10.2307/2026900

Seow, T., & Fleming, S. M. (2019). Perceptual sensitivity is modulated by what others can see. *Attention, Perception, & Psychophysics*, *81*(6), 1979-1990.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. Nature, 408(6814), 788-788.

Shteynberg, G., Hirsh, J. B., Wolf, W., Bargh, J. A., Boothby, E. B., Colman, A. M., ... & Rossignac-Milon, M. (2023). Theory of collective mind. Trends in Cognitive Sciences.

Silins, N. (2016). Cognitive penetration and the epistemology of perception. Philosophy Compass, 11(1), 24-42.

Siegel, S. (2017). The rationality of perception. Oxford University Press.

Talamanca, G. F. (2021). Joint Action without Mutual Beliefs. Kriterion–Journal of Philosophy, 35(1), 47-70.

Taylor, T. L., & Klein, R. M. (2000). Visual and motor effects in inhibition of return. Journal of Experimental Psychology: Human Perception and Performance, 26(5), 1639.

Tipples, J. (2002). Eye gaze is not unique: Automatic orienting in response to uninformative arrows. Psychonomic bulletin & review, 9(2), 314-318.

Tipples, J. (2008). Negative emotionality influences the effects of emotion on time perception. Emotion, 8(1), 127.

Tye, M. (2007). Intentionalism and the argument from no common content. Philosophical perspectives, 21, 589-613.

Zanesco, J., Tipura, E., Posada, A., Clément, F., & Pegna, A. J. (2019). Seeing is believing: Early perceptual brain processes are modified by social feedback. *Social neuroscience*, *14*(5), 519-529.

Zahavi, D. (2018). Collective intentionality and plural pre-reflective self-awareness. Journal of Social Philosophy, 49(1).