The Nature of Joint Attention: Perception and Other Minds

Lucas Battich

Graduate School of
Systemic Neurosciences

LMU Munich

Dissertation der
Graduate School of Systemic Neurosciences der
Ludwig–Maximilians–Universität München

Munich, 3rd May 2021

Supervisor: Prof. Dr. Ophelia Deroy
Chair of Philosophy of Mind
Faculty of Philosophy, Philosophy of Science and the Study of Religion
Ludwig–Maximilians–Universität München

Second Supervisor: Prof. Dr. Stephan Sellmaier
Research Center for Neurophilosophy and Ethics of Neurosciences
Faculty of Philosophy, Philosophy of Science and the Study of Religion
Ludwig–Maximilians–Universität München

Third Supervisor: Prof. Dr. med. Leonhard Schilbach
Klinik und Poliklinik für Psychiatrie und Psychotherapie
LVR-Klinikum Düsseldorf
Kliniken der Heinrich-Heine-Universität Düsseldorf

First Reviewer: Prof. Dr. Ophelia Deroy
Second Reviewer: Prof. Dr. Stephan Sellmaier

Date of Submission: 3rd May 2021
Date of Defense: 11th August 2021

Summary

From caregiver and infant playing with a toy, to singing duets or playing basketball, we frequently and effortlessly coordinate our attention with others towards a common focus. Joint attention plays a fundamental role in our social lives. It ensures that we refer to the same object, develop a shared language, understand each other's mental states, and coordinate our actions. According to some researchers, the capacity for joint attention is one of the features that distinguishes humans from other animals.

It is generally agreed that joint attention is "out in the open" among co-attenders: they are mutually aware of sharing attention to the same object or event. This mutual awareness puts the "jointness" in joint attention, and distinguishes it from cases of accidental, uncoordinated attention to the same object. How the notion of openness in joint attention should be analysed, however, is still hotly debated. Going beyond the metaphor of openness requires an examination of the mutual awareness that underlies joint attention. Additionally, current research is exclusively focused on vision. Yet we constantly coordinate attention in multisensory rich environments. Musicians jointly attend to the music they make together, we can effortlessly attend jointly to the aroma of wild spring flower, and the touch of others can help us in recognising where their attention is directed. Moreover, a narrow focus on visual joint attention may result in misleading and even prejudicial assessments of the intersubjective capacities of persons with sensory deficits (e.g., deaf, blind, and deaf-blind). A more systematic representation of how non-visual sensory resources contribute to joint attention is needed.

In this doctoral thesis, I combine philosophical and empirical methods to examine the role of perceptual experience in joint attention, including in cases involving multiple sense modalities. This aim is pursued in two related sub-objectives. First, I clarify the role of mutual awareness and perceptual experience in characterising joint attention. Second, I propose a functional framework to assess multisensory contributions to establishing and maintain-

ing joint attention and, in turn, how joint attention may affect multisensory perception.

The thesis is a collection of four individual research articles addressing these topics. In the first article, I critically examine the proposal that joint attention is based on some primitive intersubjective experiential relation, which cannot be analysed in terms of the mental states of each individual. I advance several arguments against this view and conclude that the theory is not conceptually sound. Following this work, in the second article I propose an empirically-informed account of the openness in joint attention. I suggest that mutual awareness is not something co-attenders must arrive at, but that it is often implicitly assumed. The third article proposes that joint attention is fundamentally a multisensory phenomenon, and shows in detail how non-visual senses make essential contributions to joint attention. Building on this proposal, the fourth article presents an empirical study testing whether engaging in joint attention with another person can impact one's own multisensory perceptual processing. I conclude the thesis with a general discussion of the implications of this work for philosophy, social neuroscience, and multisensory research.

Acknowledgements

There is a painting by Rembrandt, dating to 1632, traditionally known as *The Philosopher in Meditation*. It shows a minute figure in a vaulted dark interior, sitting at a table by the window. Forsaken and despondent, the philosopher travails alone, always alone. My time as a doctoral researcher could not have been more different. I have benefited from the support and generosity of many people, too many to name here.

First and foremost, I am extremely grateful to my supervisor, Ophelia Deroy, for being an inspiring and supporting mentor, for her immensely helpful intellectual and academic guidance, for all the stimulating discussions and collaborations throughout my PhD, and for very patiently enduring my often awfully written drafts and ideas. The Cognition, Values, and Behaviour group has offered me a unique and thriving environment for research between philosophy, cognitive science, and neuroscience, and I feel both enormously lucky and proud to be part of this group.

I would like to thank my second advisor, Stephan Sellmaier, for his generosity as well as the enthusiastic and rigorous discussions, and my third advisor, Leonhard Schilbach, for his support and encouragement. I'm grateful to my academic collaborators, Bart Geurts, Merle Fairhurst, Isabelle Garzorz, and Basil Wahn, for their time, insight, fruitful discussions, and mentorship. I would like to also thank the Graduate School of Systemic Neurosciences for all their support, and for creating a welcoming international and interdisciplinary community.

I benefited from many inspiring discussions with all of my colleagues at the CVBE, the Research Center for Neurophilosophy, and the Crowd Cognition group led by Bahador Bahrami. Sofia Bonicalzi, Jurgis Karpus, and Justin Sulik deserve special mention for their serious philosophical, scientific, and academic support, as well as all the non-serious banter.

Very special thanks to my fellow doctoral researchers, Anita Keshmirian, Oriane Armand, Mark Wulff Carstensen, Louis Longin, and Sofiia Rappe, all of

whom I'm happy to consider not just as my colleagues, but as close friends. Just by being there, they made my everyday PhD experience all the more rewarding. I also thank my mountain-loving GSN friends, Fabian, Stefan, and many more, for reminding me time and again that working hard also means playing hard in the Alps. Finally, special thanks go to Alice, who gracefully endured my friendship during all my academic travails.

I want to end by thanking my family, for their loving support, and for putting up with my often prolonged silence. This doctoral thesis is dedicated to three people without whom I would not be where I am now:

A Mercedes, Guillermo, y Beatriz

Contents

Chapter 1

General Introduction

In the spring of 1887, seven-year-old Helen Keller was holding her hand under the running water from a well spout. Keller had lost her sight and hearing at nineteen months old, growing into a withdrawn child and struggling to make herself understood. She lived, as she later put it, "at sea in a dense fog" — until that day in 1887, when her teacher Anne Mansfield Sullivan, herself visually impaired, taught her her first word:

> As the cool stream gushed over one hand [Miss Sullivan] spelled into the other the word *water*, first slowly, then rapidly. I stood still, my whole attention fixed upon the motions of her fingers. Suddenly I felt a misty consciousness as of something forgotten—a thrill of returning thought; and somehow the mystery of language was revealed to me. I knew then that "w-a-t-e-r" meant the wonderful cool something that was flowing over my hand. (Keller, 1903, 23)

Using their sense of touch, Sullivan made the water present to Keller through their shared attention to it, and could then name it through her movements. In what Keller later called her "soul's sudden awakening", this episode became her entry into the social world of communication and language.

Human beings are able to effortlessly engage with each other in activities that require attending to an object or event together. Engagement in joint attention provides a context for the child to associate language with its referent, contributing to vocabulary acquisition and language learning (Tomasello & Farrar, 1986; Bruner, 1998; Adamson et al., 2019). More generally, joint attention plays a key role in the development of mentalising, the ability to infer and understand other people's mental states, as well as more sophisticated

forms of social cognition and action coordination (Moore & Dunham, 1995; Carpenter et al., 1998; Eilan et al., 2005; Seemann, 2011b). According to Michael Tomasello (2008), the capacity for joint attention is one of the features that distinguishes humans from other animals.

It is generally agreed that joint attention is "out in the open" among co-attenders, where they are *mutually aware* of sharing attention to the same object or event. In this respect, joint attention differs from cases of accidental, uncoordinated attending to the same object. This mutual awareness puts the "jointness" in joint attention. As Jerome Bruner (1995) puts it, joint attention involves a "meeting of minds". How the notion of mutual awareness should be analysed, however, is still intensely debated. Additionally, current research is exclusively focused on vision. Yet we constantly coordinate attention in multisensory rich environments. Moreover, such narrow focus on visual joint attention misrepresents the intersubjective capacities of persons with sensory deficits (e.g., deaf, blind, and deaf-blind). Helen Keller's experience of language learning belies the importance of providing a principled way to explain the role of non-visual modalities in joint attention.

In this doctoral thesis, I combine philosophical and empirical methods to examine the role of perceptual experience in joint attention, including in cases involving multiple sensory modalities. This aim is pursued in two related sub-objectives. First, I clarify the role of mutual awareness and perceptual experience in characterising joint attention. Second, I propose a functional framework to assess multisensory contributions to establishing and maintaining joint attention and, in turn, how joint attention may affect multisensory perception.

The thesis is a collection of four individual research papers addressing these topics. In the first paper, I critically examine the proposal that joint attention is based on some primitive intersubjective experiential relation, which cannot be analysed in terms of the mental states of each individual (Paper I). In a paper closely following this work, I propose an empirically informed account of the openness in joint attention (Paper II). I suggest that mutual awareness is not something co-attenders must arrive at, but that it is often implicitly assumed. The third paper proposes that joint attention is fundamentally a multisensory phenomenon, and shows in detail how non-visual senses make essential contributions to joint attention (Paper III). Building on this proposal, the fourth paper presents an empirical study testing whether engaging in joint attention with another person can impact one's own multisensory processing of temporal stimuli (Paper IV). I conclude the thesis with a general discussion of the implications of this work. I outline

directions for future research in the areas of joint attention, socially shared perception, and their intersection with multisensory research.

1 The jointness of joint attention

The term *joint attention* was introduced in research on the ontogeny of communication by Jerome Bruner and colleagues (Bruner, 1974; Scaife & Bruner, 1975) to refer to infants' developing capacity to share their experiences about objects and events with others. Between the age of nine and eighteen months, most of us change from being able to guide someone's attention, through gestures or voice, to being able to reciprocally coordinate our attention with others on a third object of interest (Carpenter et al., 1998). In our everyday life, we continue to rely on this skill to communicate, share experiences, and coordinate with others. During joint attention, both co-attenders are mutually aware of each other's attention toward the same object or event. It is immediately "open" to co-attenders that they are jointly attending to the same object or event. This mutual awareness differentiates joint attention from cases where two people happen to look at the same object, unaware of the other's attention. But the notion of mutual awareness has been notoriously difficult to conceptualise. As Hannes Rakoczy writes:

> What makes such an episode one of truly joint attention? It is not sufficient that each of [the co-attenders] looks at the same target, nor that, asymmetrically, one sees the other looking somewhere and follows her gaze to the same target. [...] Rather, in some intuitive sense that conceptually proves notoriously difficult to spell out, both have to attend to the same target in joint and coordinated ways. (Rakoczy, 2018, 409)

Axel Seemann (2011a) proposes a useful distinction between reductive and non-reductive theories that attempt to explain the jointness in joint attention. *Reductive* theories explain what is to be jointly engaged towards a mutually shared object of attention by addressing the mental states of each participant separately and explaining how they come together. The reductive approach is thus "individualistic", in the sense that the collective mental activity at play in joint attention can be reductively explained in terms of coordinated individual mental activities.

One reasonable reductive approach is to treat the notion of jointness in joint attention as a special case of common knowledge or common belief and analyse it accordingly. Following Lewis (1969) and Schiffer (1972), common

knowledge is traditionally analysed with recursive propositions, so that two people have common knowledge that *p* only when they all know that *p*, and they all know that they all know that *p*, and they all know that they all know that they all know that *p*, and so on, *ad infinitum*. In a similar way, two people jointly attend to an object *x*, when they both attend to *x*, and they are aware that they attend to *x*, and they are aware that they are aware that they attend to *x*, and so on. Peacocke's (2005) account comes very close to this characterisation. According to this account, co-attenders must be aware that a whole complex state of awareness exists between them. Joint attention is possible only for beings who have some way of representing one's own and others' attention and mental states, and the ability to entertain reflexive states, such as having thoughts about thoughts (Peacocke, 2005). A common criticism against this iterative approach, however, is that it requires a psychologically implausible chain of recursive mental states, especially as infants are usually assumed to start participating in joint attention before their eighteenth month of age (Eilan, 2005; Campbell, 2018; Seemann, 2019). It also does not *feel like* we have to go through multiple inferences in order to arrive at joint attention. When you are involved in a situation of joint attention, you grasp this immediately or instantaneously. It is unclear how reductive accounts based on an iteration of mental states can explain this immediacy (Calabi, 2008).

Motivated from the failure of iterative approaches, *non-reductive* theories address the challenge of the openness in joint attention by taking it as a primitive condition. Non-reductive, anti-individualistic views in this sense notably include Campbell's relational account (Campbell, 2005, 2011, 2018), and Axel Seemann's (2011c; 2019) account, which closely follows Campbell's. Other non-reductive accounts include Naomi Eilan's proposal that joint attention is grounded in primitive conscious states of "you-awareness" and "communication-as-connection" (2015), and Michael Schmitz's account of an elementary form of collective subjecthood (Schmitz, 2015). Since John Campbell's relational view is the one that most closely engages with perceptual experience, on this thesis I will carefully assess its merits (Paper I). Other non-reductive views will be treated in comparatively less detail.

Campbell's view is based on his anti-representational, *relational* theory of perception. According to the relational theory, perceptual experiences are, in part, directly constituted by the actual perceived objects in the world. Perceptual experience is a matter of there being a causal and immediate relation between a subject and a token object in the outside world, and not a relation between subject and some abstract mental representation (see, among others, Campbell, 2002; Martin, 2004; Travis, 2004). In the case of joint attention, the fact that the other person is also jointly attending to the object is said

to be a constituent part of my experience (Campbell, 2005, 2011). This perceptual experience cannot be further reduced to the individual states of each co-attender, or to mental representations (Campbell, 2018). However, there are several problems with theories that take mutual awareness as a primitive phenomenon. Peacocke argues that taking joint attention as a primitive phenomenon seems to simply embed into the notion of a constitutive co-attender the property which is to be explained: the openness of joint attention (Peacocke, 2005). The first paper in this thesis examines in detail the primitivist view on its own grounds, and advances several arguments against it. A further motivation against a primitivist non-representational view of joint attention is that it goes against much of mainstream cognitive psychology and neuroscience. Conceptualising perceptual states in terms of representations is a useful assumption in much empirical research on vision and attention (Burge, 2005). A non-representational view of joint attention does not have the resources to fully address the relation between individual attention and perception, as it is currently studied in the cognitive neurosciences, and the phenomenon of joint attention. A different approach more in line with current empirical evidence is needed.

2 Perceptual experience and the epistemic significance of joint attention

In addressing the jointness question, most philosophers have focused on the *epistemic* role of joint attention. The general assumption is that joint attention provides the epistemic basis for joint actions, communication, and for sharing knowledge about the world around us (Campbell, 2002, 2011; Peacocke, 2005; Eilan, 2005). Thus, for example, Eilan explicitly calls the question of explicating the kind of mutual awareness we find in joint attention, as a preemptive "epistemological question" (Eilan, 2005, 4). Seemann argues that joint attention has "a special kind of epistemic power": it provides co-attenders "with the kind of common knowledge about [the co-attended] objects that makes demonstrative communication possible" (Seemann, 2019, 34). John Campbell notes that "whatever else is true of it, joint attention has an 'openness' about it, [...] in virtue of which joint attention ordinarily plays a distinctive role in rational, coordinated action" (Campbell, 2011, 417).

Due to this exclusive focus on its epistemological significance, however, the debate on how to characterise the jointness of joint attention has been mostly disconnected from debates in the philosophy of perception and perceptual experience. From the point of view of perceptual experience, the question of the

openness in joint attention takes a particular shape. Perceptual experiences are enjoyed strictly by individual perceivers, each with their own perceptual standpoint. Perception is always from somewhere: the perspective of the individual perceiver. How can different observers, with their different standpoints and perspectives, perceive objects *together*? The qualitative aspects of perceptual experience cannot be strictly identical between us. We need an account that retains the individualistic nature of perceptual experience, while explaining how we come to perceive objects and events together during joint attention (Seemann, 2019). The account I propose in this thesis, therefore, takes a reductive individualist approach: the jointness in joint attention can be accounted for in terms of the individual mental states of each co-attender. But I also take insights from non-reductive accounts: the jointness of joint attention is held by each individual as a unitary representation, which need not be constructed out of prior iterative representations (Paper II). The proposed approach, I argue, shifts the epistemic question from explaining how individuals can attain a relation that justifies shared knowledge of the perceptual environment, to explaining how individuals *abstain* from defaulting to erroneous assumptions of mutual awareness.

In virtue of its epistemic role, joint attention is arguably the foundation for the concept of a shared objective world (Davidson, 1999; Campbell, 2011; Seemann, 2019). Examining the notion of a shared openness between co-attenders may thus contribute to philosophical views on the emergence of this shared objectivity (Carey, 2009; Burge, 2010). Moreover, things being "out in the open" or "transparent" is thought to be a distinctive mark of common knowledge, common ground, and of successful communication (Stalnaker, 2002; Grice, 1957; Campbell, 2018). As one of the social phenomena with this property, elucidating in what sense the state of joint attention is "out in the open" will constitute an advancement for our understanding of more complex social and communicative interactions.

3 Multisensory perception and joint attention

Research on joint attention within philosophical and psychological disciplines has been mostly focused on the visual domain. In everyday interactions, however, many joint attention scenarios will crucially involve non-visual means of capturing someone's attention as well as non-visual targets of attention, which recruit the selection and integration of information from different senses. From singing a simple duet to playing in a symphony orchestra, musicians jointly attend to the music they make together; hunters can

jointly attend to the sound of birds in the trees, and decide how to position themselves for capture; we can effortlessly attend jointly to the aroma of wild spring flower, and the touch of others can help us in recognising where their attention is directed.

While most researchers would acknowledge that joint attention is often achieved by using multimodal cues, current measures and operationalisations of joint attention are still exclusively based on vision (Akhtar & Gernsbacher, 2008; Botero, 2016). As a consequence, there is currently no principled account of how non-visual modalities are involved in establishing and maintaining joint attention. An overemphasis on gaze coordination alone can even lead to biased assessments of an individual's ability to coordinate and interact with others. In one particularly extreme example, this vision-centric bias led to the claim that blind children must be trained in eye contact and correct facial orientation, since "the facial orientation of a blind person toward the speaker provides at least some indication [to the speaker] of [the blind person's] listening behavior" (Foxx, 1977). Research on social robotics has similarly been mostly focused on vision. Giving robots and avatars human-like eyes that would respond to ours is given priority in the design of interactive artificial agents (Admoni & Scassellati, 2017; Yang et al., 2018). But is vision the hallmark of human social cognition? Which role do the other senses play in joint attention?

To date, a few studies have started to examine auditory and tactile cues accompanying episodes of joint attention, and the strategies used by blind, deaf, and deaf-blind individuals to coordinate attention (e.g. Yu & Smith, 2013; Núñez, 2014; Depowski et al., 2015; Suarez-Rivera et al., 2019). Yet, at present, there is no clear framework to encompass other modalities and their interaction with what we know about vision in the study of joint attention. As part of this thesis, I propose that joint attention is fundamentally a multisensory phenomenon, and analyse in detail how the combination of multiple senses not only facilitates visual coordination but is even necessary for certain uses of joint attention (Paper III). A multisensory approach to joint attention brings together two areas of research across experimental psychology, cognitive neuroscience, and philosophy that are usually independent: social cognition and multisensory research. Although vision-centred research will certainly continue to provide valuable insight into the workings of joint attention, taking into account the several roles of non-visual senses will advance our knowledge of how people naturally establish, maintain and tune joint attention to a range of sensory objects and features.

A multisensory approach to joint attention will also allow distinguishing possible interactions between multisensory processes and coordinated atten-

tion. While the use of multiple sense modalities can affect how joint attention is established and maintained, engaging in joint attention may in turn affect an *individual*'s multisensory processes. The functional significance of joint attention, at least when it comes to adults, has been usually characterised at the group level: joint attention forms the basis for joint action, collective intentions, and provides a rational foundation for communication and for sharing knowledge about our environment.

But a recent body of research is emerging that shows the influence of socially coordinated attention on individual perceptual and cognitive processes, including processes which are not per se typically seen as part of social cognition, such as mental rotation, memory, and perceptual sensitivity (Mundy, 2018; Becchio et al., 2008; Shteynberg, 2015). For example, experiencing joint attention with another person facilitates the detection and discrimination of visual objects (Frischen et al., 2007), and it enhances a participant's mental spatial rotation performance when judging the handedness (left or right) of images of hands at different angles (Böckler et al., 2011). Engaging in joint attention to a common object facilitates information encoding in working memory for that object (Kim & Mundy, 2012; Gregory & Jackson, 2017), and impacts the affective appraisals of objects in the environment (Bayliss et al., 2006). Participants are also better at detecting nearly imperceptible patterns when someone else is also looking towards the same patterns (Seow & Fleming, 2019). These studies, however, are predominantly based on visual perceptual targets. But does joint attention affect an individual's *multisensory* perception?

One prevalent hypothesis regarding the functional role of joint attention is that it deepens or enhances the encoding of stimulus information in ways that are not observed when information is individually attended (Becchio et al., 2008; Mundy, 2016, 2018). Joint attention adds a further level of selectivity to an individual's attention. An open question is whether this hypothesis extends to the temporal multisensory processing of events. Although theoretically motivated, this is an empirical question not amenable to philosophical theorising. To address this question, therefore, I conducted a behavioural psychophysics study measuring how engaging in joint attention with another person can impact one's own multisensory perceptual processing (Paper IV). We used the sound-induced flash illusions, which are reliable indicators of temporal multisensory integration (Shams et al., 2002; Andersen et al., 2004; Keil, 2020; Hirst et al., 2020). In the fission illusion, a single flash accompanied by two task-irrelevant auditory beeps induces a visual percept of two flashes; in the fusion illusion, two flashes are perceived as one when accompanied by one beep.

Previous studies on the impact of joint attention on spatial multisensory processing have focused on the effect of social eye gaze cues, elicited through an artificial partner or avatar (De Jong & Dijkerman, 2019; Nuku & Bekkering, 2010). As noted above, however, joint attention involves the minimal understanding that one is currently sharing attention to the same object or event with another agent. To take this interpersonal aspect into account, we tested pairs of participants, who performed the flash-counting task either alone, or in pairs sitting in close proximity. Following the hypothesis that engaging in joint attention enhances the relative processing of the jointly attended visual target, we expected a change in the relative weight accorded to visual and auditory information, so that the influence of the auditory distractor, and thus of the strength of both illusions, will be reduced.

We replicated the effect of both sound-induced flash illusions, as measured by the number of flashes reported. We did not find any statistically significant effects for our main hypothesis. People did not perform better nor worse across the different social conditions tested, for both fission and fusion illusions. Using signal detection measures, we found that people's criterion bias was less affected by the auditory beeps for the fusion illusion (their bias decreased), when engaged in joint attention as contrasted with the individual condition, although this reduction was not enough to effect a significant change in the mean number of flashes reported. Our findings indicate that the strength of both fission and fusion illusions is not sensitive to engagement in joint attention. Our results highlight the limitations of the hypothesis that joint attention enhances stimulus information encoding and processing in multisensory settings. With this study, we provide grounds for future work in whether and how social factors may influence multisensory processing.

4 Thesis synopsis

This thesis is concerned with the role of perceptual experience in joint attention, and aims to contribute to two questions:

- What is the jointness in joint attention, and how it should be analysed?
- How different senses shape joint attention and, conversely, how can joint attention affect perception across modalities?

The dissertation is made of a collection of individual research papers addressing these questions. In the course of these papers, I advance, first, an empirically-informed account of joint attention and its basis on mutual

awareness. Second, I show that non-visual senses make essential contributions to joint attention, and develop experimental methods for testing how joint attention can affect perception across sense modalities. Below, I include summaries of each research paper, and information about where they have been accepted for publication, if applicable.

Paper I: Joint attention and perceptual experiences

Joint attention refers to the coordinated focus of attention between two or more individuals on a common object or event, where it is mutually "open" to all attenders that they are so engaged. But there is no consensus on how to analyse this mutual "openness". One prominent account, based on relational or "naive" theories of perception, is to view joint attention as a primitive relation of consciousness, which is not to be explained in terms of the mental states of each individual (Campbell, 2011, 2018). In this paper, we critically assess this approach and find it conceptually unsound. This approach to joint attention has proved to be attractive to both philosophers and scientists concerned with social cognition and its development (e.g., Moll & Meltzoff, 2011). That this view is not conceptually feasible is therefore critical to several lines of research in social cognitive psychology and philosophy of mind. Our arguments have wider implications for debates in theories of common knowledge, demonstrative communication, and the philosophy of perception.

This paper has been published under open access (Creative Commons Attribution 4.0 International License) as:

Battich, L. and Geurts, B. (2020). Joint attention and perceptual experience. *Synthese*, doi: 10.1007/s11229-020-02602-6.

Paper II: Opening up the openness of joint attention

Joint attention is often defined as a mutually "open" relation between co-attenders: they are mutually aware of being so engaged. But how should this openness be characterised? In this paper, I first distinguish between two explanatory aims often conflated in the current debate: the aim of explaining the normative role of joint attention in justifying joint endeavours and shared knowledge of the world, and the cognitive aim of explicating the psychological capacities and wherewithal involved in joint attention. I argue that current theoretical accounts of joint attention are primarily designed to tackle the normative concerns, and their problems arise when they conflate these concerns with psychological ones. Drawing from evidence in developmental

and cognitive psychology, I outline the case for a cognitive account of joint attention based on a weaker notion of openness and mutual awareness. My arguments have direct relevance to philosophical debates on shared mental states and, in particular, the notion of a common ground of knowledge shared between interactants.

Paper III: Coordinating attention requires coordinated senses

Playing tennis, singing together, ordering a cake: we effortlessly coordinate each other's attention towards a common focus in rich multisensory ways. In this paper, we propose that joint attention is fundamentally a multisensory phenomenon. We highlight that joint attention relies on the strategic coordination of many senses and not just from following other people's eye gaze or pointing gestures, and propose a novel framework to assess the multisensory contributions to joint attention. Our paper bridges two research areas: the study of joint attention, which is embedded in the fields of social cognition, developmental psychology and social robotics, and the study of multisensory attention within psychology and cognitive neuroscience. The former remains primarily focused on vision and can benefit from research on the role of joint attention across sensory modalities. Multisensory research, on the other hand, remains centred on single individuals. As most everyday objects and events that we jointly attend to are multisensory, whether and how social factors influence multisensory processing opens new avenues of research. We outline clear directions for future experimental research, and detail the implications for social robots, clinical diagnostics, and theoretical debates on shared objectivity.

Battich, L., Fairhurst, M., and Deroy, O. (2020). Coordinating attention requires coordinated senses. *Psychonomic Bulletin & Review, 27*(6), 1126–1138. doi: 10.3758/s13423-020-01766-z.

Paper IV: The impact of joint attention on the sound-induced flash illusions

Among the open questions in multisensory research is whether social factors influence multisensory processing. In this pre-registered study, we investigated whether joint attention impacts temporal multisensory integration. A leading hypothesis on the functional role of joint attention is that it enhances

the encoding and processing of the jointly attended perceptual information (Mundy, 2018). We tested whether this hypothesis holds for temporal multisensory processing by using the well-documented sound-induced flash illusions, where a single flash is perceived as two when accompanied by two auditory beeps (fission), and two flashes as one when accompanied by one beep (fusion) (Shams et al., 2002; Andersen et al., 2004). We compared participants' performance in a flash-counting task in three conditions: alone, jointly attending with someone else, and a non-joint attention social condition. If the processing of the jointly attended visual target is facilitated relative to the sound distractor, then we would expect that the illusions would be reduced. We found that people's criterion bias in the fusion illusion diminished when they engaged in joint attention. Importantly, however, as measured by the number of flashes reported, people did not perform statistically better nor worse across the different social conditions tested, indicating that the strength of the illusions is not sensitive to engagement in joint attention. These findings show the limitations of the hypothesis that joint attention enhances stimulus information processing in multisensory settings. This study provides grounds for future work in studying the effects of joint attention on different multisensory processes.

This paper is has been accepted for future publication at *Attention, Perception & Psychophysics*:

Battich, L., Garzorz, I., Wahn, B., and Deroy, O. (forthcoming 2021). The impact of joint attention on the sound-induced flash illusions. *Attention, Perception and Psychophysics*.

Chapter 2

Paper I: Joint Attention and Perceptual Experience

Battich, L. and Geurts, B. (2020). Joint attention and perceptual experience. *Synthese*, doi: 10.1007/s11229-020-02602-6.

Author contributions:

L.B. conceived of the research idea and arguments presented in the paper, wrote the original draft, and revised the paper with help from B.G.

Synthese
https://doi.org/10.1007/s11229-020-02602-6

Joint attention and perceptual experience

Lucas Battich[1,2] · **Bart Geurts**[3,4]

Received: 24 October 2018 / Accepted: 22 February 2020

Abstract
Joint attention customarily refers to the coordinated focus of attention between two or more individuals on a common object or event, where it is mutually "open" to all attenders that they are so engaged. We identify two broad approaches to analyse joint attention, one in terms of cognitive notions like common knowledge and common awareness, and one according to which joint attention is fundamentally a primitive phenomenon of sensory experience. John Campbell's relational theory is a prominent representative of the latter approach, and the main focus of this paper. We argue that Campbell's theory is problematic for a variety of reasons, through which runs a common thread: most of the problems that the theory is faced with arise from the relational view of perception that he endorses, and, more generally, they suggest that perceptual experience is not sufficient for an analysis of joint attention.

Keywords Joint attention · Perceptual experience · Common knowledge · Relationalism · Perception · John Campbell

1 Introduction

Unbeknownst to each other, we are looking at the same piece of cake. Our attention is shared, but we don't know that it is, and therefore the fact that our attention is shared doesn't affect us in any way. But now we both come to realize that we are looking at the same piece of cake. Our visual attention becomes coordinated: alternating glances between the cake and the other, each is now aware of the other's attention. Everything

Lucas Battich
lucas.battich@campus.lmu.de

1 Faculty of Philosophy, Ludwig-Maximilian-University Munich, Geschwister-Scholl-Platz 1, 80359 Munich, Germany

2 Graduate School of Systemic Neurosciences, Ludwig-Maximilian-University Munich, Munich, Germany

3 Radboud University, Nijmegen, The Netherlands

4 HSE University, Moscow, Russian Federation

Published online: 05 March 2020

about our attention is out in the open, and if one of us were to say, "It's mine!", the intended referent would be obvious. Thus we went from a state of (merely) shared attention to joint attention. But what has changed? Or, as Hobson's title (2005) has it, "What puts the jointness into joint attention?"

In recent years, the jointness question has received quite a lot of attention. It is agreed that joint attention is a ubiquitous phenomenon, and that it is important to human social interaction, because it helps us to coordinate our actions and beliefs. Since the term was introduced in developmental research by Jerome Bruner and colleagues (Bruner 1974; Scaife and Bruner 1975), joint attention has been considered a milestone in children's social and cognitive development (Moore and Dunham 1995; Carpenter et al. 1998; Adamson et al. 2019), and shortcomings in joint attention have been associated with the onset of autism spectrum disorders (Hobson and Hobson 2011; Mundy 2016).

But there is no consensus on what joint attention *is*. The starting point for many (though not all) authors is that the jointness of joint attention is "open" between both attenders. It is fully and immediately transparent to them that they are jointly attending to the same object or state of affairs (thus, e.g.,Bakeman and Adamson 1984; Tomasello 1995; Peacocke 2005; Calabi 2008; Campbell 2011; Carpenter and Liebal 2011; Eilan 2015). The challenge is to go beyond the metaphor of openness.

One view is that joint attention is to be understood in terms of common knowledge, or some related notion like common awareness, common belief, common acceptance, etc. There are various ways of fleshing out this idea, but the simplest is to define that we jointly attend to an object iff it is common knowledge between us that each of us is attending to it. Assuming that common knowledge can be defined in terms of knowledge (or a related epistemic notion like belief or awareness), this proposal is reductionist in the sense that it defines joint attention in terms of individual mental states.

An alternative to the knowledge-based approach is to view joint attention as a primitive relation, which is irreducible to the individual states of its relata (e.g.,Calabi 2008; Seemann 2004). John Campbell's (2005, 2011, 2018) "relational" theory is a prominent representative of this view. On Campbell's account, when jointly attending to the cake, each of us *experiences* the other as jointly attending to the cake, such that you are a constituent of my visual experience, as I am a constituent of yours. Whatever epistemic significance joint attention has in coordinating our actions and beliefs, it results from this sensory experiential character. Unlike the knowledge-based approach, Campbell's analysis is based on the idea that joint attention is "fundamentally a phenomenon of sensory experience" (2011, p. 415) and seeks to avoid referring to judgments, inferences, appeals to knowledge, beliefs, or any other higher cognitive processes. For this reason, Campbell's style of analysis has proved to be attractive to cognitive and developmental psychologists (e.g.,Moll and Meltzoff 2011; Hobson and Hobson 2011). Moreover, the relational view has been claimed to support and complement an interactionist approach to social cognition based on embodied, embedded, and extended interactive processes (Gallagher 2010; León et al. 2019). In these and other ways, the feasibility of the relational approach is critical to several lines of thinking in social cognitive psychology and the philosophy of mind and language.

We argue that Campbell's relational account of joint attention fails to deliver on its promises. In doing so, we do not wish to advocate the knowledge-based approach, let

alone defend it against objections from Campbell or other critics. Rather, we intend to assess the merits of the relational approach in its own right, and will argue that, at several points, Campbell's theory threatens to collapse into its competitor. Therefore, we will need to discuss the knowledge-based view in some detail.

To begin with, Sect. 2 elaborates on the knowledge-based approach that serves as the foil to Campbell's relational account, which is presented in Sect. 3. Campbell's account is underdeveloped in several respects, and therefore we will need to consider various ways of making it more precise. We then argue that the relational definition of joint attention either results in an infinite regress of perceptual states or requires a construal of the notion of "co-attention" that is substantially identical with the notion of "normality" employed by knowledge-based theories, which, according to Campbell, shouldn't be part of an explanation of joint attention (Sect. 4). Finally, we discuss two further issues having to do with attention monitoring (Sect. 5) and failures of joint attention (Sect. 6).

The recurring theme throughout our discussion is that the problems which Campbell's theory runs into are due to tensions in his claim that joint attention is "fundamentally a phenomenon of sensory experience", and therefore not to be explained in terms of knowledge, belief, or awareness (2011, p. 415, 2018, p. 120). While joint attention undeniably involves sensory experience, our discussion suggests that an explanation of the phenomenon will have to factor in at least some knowledge, belief, or awareness.

2 The knowledge-based approach

Campbell presents his theory of joint attention as a superior alternative to the knowledge-based approach. On the latter view, joint attention can be defined as follows:[1]

– A and B are jointly attending to x iff it is common knowledge between A and B that each of them is attending to x.

It doesn't matter for our purposes whether this particular definition is the best way of dealing with joint attention in terms of common knowledge; nor does it matter whether, e.g., common awareness, common belief, or common acceptance might be preferable to common knowledge. The only thing that matters is that all analyses that take this general approach have two features in common: they refer to cognitive states and they entail that these states give rise to iterative structures like the following:

– *p* is common knowledge between A and B iff A knows that *p*, B knows that *p*, A knows that B knows that *p*, B knows that A knows that *p*, and so on *ad infinitum*.

Structures like this are the fingerprint of common knowledge, common belief, and so on (Lewis 1969; Schiffer 1972; Geurts 2019). On a knowledge-based account, it is this iterative structure that is held to capture the jointness in joint attention.

[1] Throughout this paper, we confine our attention to instances of joint attention and of common knowledge involving two participants.

It is important to be clear about the status of this iterative structure, for it is often misunderstood as implying that A and B cannot have common knowledge unless (*i*) they make an infinite number of inferences and (*ii*) they mentally represent the outcomes of all these inferences. The first misunderstanding was addressed by David Lewis (1969, p. 53) even before it arose: "Note that this is a chain of implications, not steps in anyone's actual reasoning. Therefore there is nothing improper about its infinite length." While this may not help very much to explain what common knowledge is, it should have sufficed to dispel the notion that it requires an infinite number of inferences as a precondition. Regrettably, Lewis's remark was widely ignored in the subsequent literature, but that is as it may be: objection (*i*) is merely a denial of Lewis's remark, and now that it has been dispensed with, objection (*ii*) falls by the wayside, too.

Having characterized common knowledge in terms of the iterative structure shown above, it remains to be seen how common knowledge is achieved. Following Lewis (1969) and Schiffer (1972), it is generally accepted that there are many types of finite situations, or "bases" as Lewis calls them, that generate common knowledge. People interacting with each other will soon find out that they share the same language, the same social background, the same hobbies, and so on, and any of these commonalities will serve as a basis for common knowledge. In the case of joint attention, part of the relevant basis will be that A and B take each other to know "that if a 'normal' person (i.e. a person with normal sense faculties, intelligence, and experience) has his eyes open and his head facing an object of a certain size (etc.), then that person will see that an object of a certain sort is before him" (Schiffer 1972, p. 31). If this normality condition is not fulfilled, joint attention cannot be achieved. For example, if A knows that B's low eye pressure seriously distorts her vision, then A and B cannot jointly attend to their cake.

The knowledge-based approach invites (but does not entail) the hypothesis that, apart from the normality condition, joint attention may be affected by practically any kind of background knowledge. Just as individual attention is driven by goals, intentions, and beliefs, so is joint attention. Tomasello (2014) notes that in joint attention I must be sensitive to the features of an object or situation that are relevant for you. Just following your line of gaze is not enough (cf. Moll and Tomasello 2007). To illustrate, consider the following scenario. If you point at a tree, I can follow your pointing gesture, but that does not tell me whether you are attending to the apples it carries, its smooth trunk, or the fungus on its bark. Perhaps we have been out foraging for apples or, alternatively, you have been tasked with the care of fungus-infected trees. My perceptual experience will be the same in both scenarios, since our lines of sight converge on the same tree. It is our shared background knowledge that enables me to determine which aspects of the visual scene you are focusing on, and to engage in joint attention with you.

The knowledge-based view is consistent with a range of positions on how much and what kinds of background knowledge are required for joint attention. However, since Campbell restricts his attention to knowledge-based theories that adopt the normality condition, we will make the same restriction here.

Campbell (2005) criticizes the knowledge-based approach for being cognitively demanding and psychologically unrealistic, because it requires infinitely many inferences and infinitely many levels of mental representation. The same objection is made

by Eilan (2005) and Carpenter and Liebal (2011), among others, and Calabi (2008) goes so far as to suggest that the theory entails that joint attention requires a grasp of the concept of infinity. As we have already seen, we maintain that this line of criticism is a spurious one, and we will consider it no further. Nevertheless, and regardless of how the knowledge-based approach fares, one can of course maintain that the relational theory presents a viable alternative. We now turn to that account.

3 The relational theory of joint attention

On an orthodox analysis, perceptual experiences are constituted not only by our surroundings, but also by our mental representations. Campbell's theory of joint attention extends his anti-representational theory of perception, which he calls "relational". On the relational view, "the phenomenal character of your experience, as you look around the room, is constituted by the actual layout of the room itself: which particular objects are there, their intrinsic properties, such as colour and shape, and how they are arranged in relation to one another and to you" (Campbell 2002, p. 116; see also Martin 2004; Travis 2004; Crane 2006). On this view, perception is not a matter of representing objects, but involves a non-representational relation between the perceiver and the token object perceived. Up to a point, this agrees with our intuitions. When we look around the room, we would normally say that we experience the room itself and its contents. Therefore, Campbell's account is a representative of what is sometimes called a "naïve realist" view on perceptual experience.

Campbell construes perception as a three-place relation: "S perceives x as being F", where the F-term stands for something like the aspect under which x is perceived. If I look around the room, I see the objects it contains as having certain intrinsic properties and being arranged in certain ways relative to one another and to myself.[2] Campbell doesn't discuss F-properties and -relations in any detail, which is unfortunate, because they play a key role in his account of joint attention, as we will see. However, for our purposes it suffices to note that, as defined by Campbell, perception is a purely extensional relation, which entails that, if F and G are the same, "S perceives x as being F" is equivalent to "S perceives x as being G".

Campbell considers perception to be a primitive relation, in the sense that it is not to be analyzed in such terms as "x causes S to have a representational content as of something being F", where S experiences this representational content (Campbell 2002, pp. 117-18). More generally, perception is a relation between subjects, objects, properties, and relations that is irreducible to other mental states.[3]

Campbell's analysis of joint attention builds on his relational account of perception:

> On a relational view, joint attention is a primitive phenomenon of consciousness. Just as the object you see can be a constituent of your experience, so too it can

[2] This is Campbell's (2002) analysis. In later work, he introduces "standpoints" as a further ingredient of perception (2009). Although it may be that at least some F-relations can be subsumed under this notion, it seems to us that Campbell's views on "perceiving as" remain unaffected by its addition.

[3] For critical assessments of the relational view of perception, see Burge (2005), Byrne and Logue (2008) and Nanay (2014). For attempts to reconcile the relational and representational views, see Nanay (2015) and De Sá Pereira (2016).

> be a constituent of your experience that the other person is, with you, jointly attending to the object. This is not to say that in a case of joint attention, the other person will be an object of your attention. On the contrary, it is only the object that you are attending to. It is rather that, when there is another person with whom you are jointly attending to the thing, the existence of that other person enters into the individuation of your experience. The other person is there, as co-attender, in the periphery of your experience. (Campbell 2005, p. 288)

On this account, joint attention involves an object x and two individuals who experience each other as co-attending. That is to say:

- A and B are jointly attending to x iff
  - A perceives x as being co-attended by B, and
  - B perceives x as being co-attended by A.

Here the F-term of relational perception à la Campbell is instantiated with the property of being co-attended by the other. Thus, if I am alone eyeing the cake on the table, and you arrive to engage in joint attention with me, then there is a change in my perception of the cake: I now *see* it as being co-attended to by you.

But what does it mean for two people to co-attend to an object? Campbell doesn't say, but the most natural answer, it seems to us, is that co-attention is just a variant expression for joint attention: A and B perceive an object as being co-attended by the other iff they jointly attend to it. Alternatively, and perhaps less likely, co-attention and joint attention may be distinct concepts. We will consider both options shortly.

According to Campbell, joint attention differs from other forms of simultaneous attention in two respects. If A and B are jointly attending to x, then first, A and B monitor each other's attention, and second, A's attention is one of the factors controlling B's attention, and vice versa (cf.Tomasello 1995, p. 107). In line with his relational principles, Campbell sees attention monitoring and control as relations that are to be fleshed out in causal terms. For example, Campbell stipulates that, in order for joint attention to be achieved, B's continued attention to x must be one of the causal factors for A's continuing to attend to x, and vice versa, A's continued attention to x must be one of the causal factors for B's continuing to attend to x (Campbell 2005, p. 289).

Campbell assumes that "this coordination of attention may involve the use of subpersonal mechanisms, rather than explicit, personal-level thoughts about the direction of the other person's attention" (2005, p. 288). Here the personal/subpersonal distinction coincides with the introspectable/non-introspectable distinction, and personal mental states are taken to include sensations, emotions, beliefs, desires, and rational and deliberate thinking. The subpersonal processes for monitoring and control cannot account for the jointness of joint attention, because joint attention is a personal-level phenomenon and "it is hard to see what [these subpersonal processes] contribute to the subject's psychological life" (Campbell 2011, p. 416). The jointness in joint attention is a personal-level state. This is why Campbell seeks to explain joint attention in terms of the conscious experience of having the other as a co-attender. On Campbell's view, it is precisely in virtue of its perceptual experiential character that joint attention can have the epistemic significance that it has.

Campbell maintains that his relational theory of joint attention is free of the difficulties that he and others associate with the knowledge-based approach. First, it doesn't involve the infinitely iterating structures that are the hallmark of knowledge-based theories. Second, it doesn't appeal to background knowledge, and in particular, it doesn't appeal to anything like the normality condition, which on a knowledge-based account is instrumental in generating these iterative structures. In short, on a relational analysis, joint attention is defined in terms of perceptual experience (Campbell 2011, p. 415). In the following sections, we argue that this position is untenable.

4 Co-attention

As a matter of logical necessity, co-attention and joint attention are either the same thing or not. Peacocke has noted that, if co-attention and joint attention are the same thing, the notion of a co-attender presupposes the property which is to be explained, i.e. the openness of joint attention (2005, p. 300). Nevertheless, Campbell's own discussion suggests rather strongly that, for him, joint attention and co-attention are identical: to be a co-attender is just to stand in the primitive three-place experiential relation, with another co-attender, to a common object (2011, p. 420). Since joint attention is an extensional relation, this entails that, if $x = x'$, then A and B are jointly attending to x if and only if they are jointly attending to x' (2011, p. 424). Thus, it follows that, if A and B jointly attend to x:

- A perceives x as being co-attended by B,
- B perceives x as being co-attended by A,
- A perceives x as being perceived by B as being co-attended by A,
- B perceives x as being perceived by A as being co-attended by B,
- A perceives x as being perceived by B as being perceived by A as being co-attended by B,
- B perceives x as being perceived by A as being perceived by B as being co-attended by A,
- and so on *ad infinitum*.

So now joint attention involves the same sort of infinite iterations that, according to Campbell, invalidate the knowledge-based approach. Hence, in this respect, Campbell's theory turns out to mimic the knowledge-based approach. By his own lights, this is an unwanted result, since part of the motivation for claiming that "joint attention is a primitive phenomenon of consciousness", is to avoid the complex iterations of mental states that plague the knowledge-based approach (Campbell 2005). While Campbell is not fully clear on the notion of primitiveness, it is often assumed that an infinite regress is blocked just because joint attention is a primitive relation (e.g. Calabi 2008; Seemann 2004; Eilan 2015). We fail to see how this line of defence might work. Consider the following case. In most versions of propositional logic, conjunction is a primitive, non-reductive relation, in the sense that it cannot be reduced to other relations (other versions may take disjunction or implication to be primitive instead). This doesn't prevent it from licensing endless series of inferences. For example, if p & q holds, then:

– p & p & q,
– p & p & p & q,
– p & p & p & p & q,
– and so on *ad infinitum*.

Evidently, primitiveness in itself does not impede recursion, and there is no reason to suppose that the primitiveness of joint attention will block the infinite regress pictured above. As Campbell's defines it, co-attendance may not be reducible to other individual mental states, but that doesn't prevent the definition from generating a recursive regress. What is crucial to the regress is that the relation is extensional. If primitiveness is meant to resolve the issue, we are owed a positive explanation of what it is and how it accomplishes this feat.

Thus, the assumption that co-attention is joint attention, which Campbell seems to subscribe to, leads into major trouble for his account. Therefore, let's consider the possibility that co-attention and joint attention are not the same thing, and let's grant, if only for the sake of the argument, that this will block the infinite regress that would otherwise ensue. According to Campbell, when we are engaged in joint attention there is a difference between how I am related to the co-attended object and how I am related to you. Each person is "there" and enters the other's experience, "as co-attender" (Campbell 2011, p. 419). We will not try to provide a full-dress definition of co-attention, but will merely consider what are likely to be some of the minimal conditions that must hold for someone to enter another person's perceptual experience as a co-attender.

In order to experience B as co-attender, A must be able to recognize that B co-attends to x with her. Apart from the fact that this is a natural assumption to make, it is also in line with what Campbell writes about other F-properties:

> To experience the shape of a solid object you must have some capacity to recognize manifest sameness of shape across movements by you or by the object. Otherwise it is hard to see how you could be said to be encountering the property of three-dimensional shape at all. (Campbell 2009, p. 288)

By analogy, having the capacity to recognize co-attention when one encounters it is a necessary precondition for joint attention. What does this capacity involve? For starters, a plausible candidate is the ability to recognize the other as an animate entity, separate from oneself, and to sense, however minimally, the other's agency (e.g. that they have goals different from one's own). But clearly this won't suffice for me to see you as a co-attender rather than merely as a person who happens to be looking in the same direction, for example, or who is incapable of looking in the first place. Thus we are led to suppose that the ability to recognize co-attention requires the ability to include as candidate co-attenders people whose line of sight intersects with the target object, and exclude the blind, blindfolded, comatose, and so on. But these are precisely the sort of requirements that make up the normality condition on which the knowledge-based view is based.

This is bad news for Campbell's account for two reasons. First, because he explicitly seeks to avoid any appeals to knowledge, beliefs, or awareness of the two participants (Campbell 2018, p. 120). Secondly, because on the knowledge-based view, the nor-

mality condition is the linchpin in generating the endless iterations of psychological states that Campbell rejects. Again, we have come to a point at which the relational account threatens to converge with the knowledge-based account.

The key observation on which the foregoing argument is based is just that the kind of *content* that seems to be needed to flesh out the notion of co-attention is the same as what goes into the normality condition. No assumptions have been made about the *nature* of that content, except for the fact that, if this content is a prerequisite for my joint attentional perceptual experience, it cannot itself be accounted for in terms of that experience. This point bears emphasizing because the normality condition has been held to require a grasp of a concept of psychological normality and all that goes with it (e.g., Peacocke 2005). As far as we can tell, that is not the case. Whatever kind of content is involved in co-attention will work for normality, too (cf. Calabi 2008).

As discussed in Sect. 2, while the knowledge-based view is consistent with the hypothesis that, in principle, joint attention may be affected by any kind of knowledge, it allows for a range of positions on how much and what kinds of knowledge are required for joint attention; the normality condition may be seen as an attempt at capturing at least some of that knowledge. In the foregoing we were led to conclude that, whatever co-attention may be, it seems likely to involve the same kind of knowledge. The bottom line is that *at least some* knowledge must be involved in any analysis joint attention. If we try to do without any form of knowledge whatsoever, a feasible account of joint attention is outside our reach. Therefore, the relational view is on the wrong track. In the remainder of this paper we discuss two further issues that reinforce this conclusion.

5 Causal monitoring

On Campbell's view, when we are engaged in joint attention, you are a constitutive part of my experience. For this to happen, some causal conditions must be met (Campbell 2005, p. 288). Part of your causal contribution to my experience is that you are continuously attending to the object that I'm attending to. More formally:

- A's continued attention to x must be one of the factors causally sustaining B's continuing to attend to x, and
- B's continued attention to x must be one of the factors causally sustaining A's continuing to attend to x.

These conditions can be interpreted strongly or weakly. On a strong interpretation, causal monitoring must be literally continuous, i.e. uninterrupted. This interpretation is suggested by Campbell's (2005, p. 289) own words, which we have reproduced almost verbatim. On the weak interpretation, monitoring need not be continuous in order to sustain joint attention. It is not hard to see that, in both versions, Campbell's causal conception of monitoring is problematic: on the strong interpretation it is unrealistic, and on the weak interpretation it is hard to see how it could be causal. Since the two interpretations are jointly exhaustive, it is doubtful that the mutual monitoring on which joint attention is generally agreed to be based is a causal relation.

Consider the strong version first. On this interpretation, you have to keep looking at our cake without interruption in order for our state of joint attention to persist. If you divert your gaze even for a second, the causal connection is broken, I cease to experience you as a co-attender, and our joint attention is no more. This is clearly wrong. When we jointly attend to our cake, for example, we typically alternate gazes between the cake and each other; if both of us were fixedly staring at the cake without checking with each other every once in a while, we wouldn't be engaged in joint attention. Hence, on a strong interpretation, causal monitoring fails to account for the facts.

On a weak interpretation, your eye gaze is allowed to shift between the cake and myself (and perhaps other objects as well). But then how can we account in purely causal terms for the difference between a solid 10-minute bout of joint attention and an interval of the same length during which our joint attention is briefly interrupted every now and then? On a knowledge-based model, joint attention may be sustained, in part, by informational processes that enable us to distinguish between genuine interruptions and merely apparent ones. For example, if it is common knowledge between A and B that each is equally interested in x and the other, then alternating gazes between each other and x are more likely to be experienced as joint attention than if it is common ground that x is of predominant interest for both A and B. It is hard to see how such observations could be accommodated by a purely causal model and the perceptual relation it is meant to support. To explore this point a bit further, we turn to our last topic: failures of joint attention.

6 When joint attention fails

Once again, we have been jointly attending to our cake for a while, when you start daydreaming about your next holiday, and although you're still gazing at the cake, your mind is now elsewhere. Hence, our episode of joint attention has come to an end, but as far as I'm concerned we are still looking at the cake together. How is this possible? In his 2011 article Campbell diagnoses the situation as follows:

> Being an experiential relation, like "___ sees ___", it is introspectable: X can tell just by reflection that he or she is co-attending with Y to Z. However, here as so often, introspection is not an infallible source of knowledge. You may think you are co-attending with Y to Z even though Y left long ago. (Campbell 2011, p. 419)

Based on introspection, I believe that we are jointly attending to the cake while in fact we're doing no such thing anymore. On Campbell's account, this is a scenario that theories of direct perceptual experience are all too familiar with. Consider the following experiment. A subject is looking at a tennis ball which, during the 200 milliseconds of an eye blink, is replaced with another, qualitatively indistinguishable ball. So our subject doesn't notice the change, and as far as she is concerned her perceptual experience is the same as before. For a relational theorist like Campbell the case is clear cut: the replacement causes the subject to enter a new perceptual state, even if she fails to notice it (cf. Martin 2004; Schellenberg 2010).

This claim is controversial, but it is the logical consequence of the premise that, in veridical perception, external objects and their properties "partly constitute one's conscious experience" (Martin 1997, p. 83). This premise clashes with the intuition that two conscious perceptual experiences that are indistinguishable for a subject are necessarily the same (Martin 1997, p. 81). It is generally agreed that these two notions are difficult if not impossible to square. However, we will not address that issue here, and merely want to point out that, compared to the tennis ball experiment, cases of false joint attention raise additional issues for the relational account.

First, whereas in the tennis ball experiment the perceived objects are numerically distinct, in our case of failing joint attention it is just the fact that you cease to pay attention to the cake that, according to Campbell, causes a change in my perceptual experience. By hypothesis, there are no external factors that might causally account for my change of perceptual state. Only *neural* changes in you might conceivably qualify for this job. Therefore, Campbell owes us an account of how covert changes in the brain states of one person can affect perceptual experiences in another.

Secondly, since the relational view allows for dissociations between my perceptual experience and my beliefs about my perceptual experience, it also allows for the possibility that I am engaged in joint attention but mistakenly believe that I am not. But this seems to be at odds with the key feature of joint attention that, as noted in the introduction, all parties agree on: joint attention is public, it has a special kind of openness or mutual manifestness. On Campbell's account, this openness is constituted by my perceptual experience and yours, and therefore I can mistakenly believe that we are not engaged in joint attention because I am wrong about my experience. This sounds like a downright contradiction to us. It is one thing to suppose that I can wrongly believe that we *are* engaged in joint attention; this is a possibility that every theory should allow for. But it is quite another thing to suppose that I can wrongly believe that we *are not* engaged in joint attention. This is a possibility that, in our view, should be ruled out by the very notion of joint attention, and if this much is true, it is problematic that Campbell's account fails to do so.

7 Conclusion

The relational approach has been touted as a superior alternative to the knowledge-based approach, and has been claimed to provide an account of joint attention anchored in its perceptual experiential character, which avoids an infinite regress of inferences, does not require conceptual understanding, and generally imposes minimal demands on processing and representation. For these reasons, it has proved to be attractive to philosophers and psychologists concerned with social cognition and its development.

In the foregoing we have argued that Campbell's theory is untenable, and at several points comes perilously close to collapsing into its knowledge-based competitor. To begin with, the relational analysis either results in an infinite regress of psychological states or necessitates elaborations of the notion of "co-attention" that make it indistinguishable from the notion of "normality" employed by knowledge-based theories. Further, the theory requires a causal notion of attention monitoring which is either too strict to be realistic or so loose that it cannot be a purely causal notion in the

first place. Finally, the theory implies a counter-intuitive dissociation between my perceptual experience and my beliefs about my perceptual experience, which becomes problematic when considering cases of joint-attention failure.

Where do we go from here? The relational view is predicated on the assumption that joint attention is fundamentally a type of primitive perceptual state, not itself susceptible to explanation in terms of the knowledge, beliefs, or awareness of the two participants (Campbell 2018, p. 120). Our discussion suggests that the difficulties the relational view faces may be tackled by abandoning this assumption, and by taking into account the knowledge, beliefs, or informational states of each participant in joint attention. Of course, joint attention certainly includes perceptual experiential aspects. Our discussion of the relational view suggests, however, that these perceptual aspects are not sufficient for an account of joint attention. Further work in a theoretical explanation of joint attention may do well by taking into account the combination of sensory experience and individual epistemic states.

To sum up, we believe that our arguments raise serious issues for theories of joint attention that adopt a relational take on perceptual experience. More generally, they suggest that theories anchored in perceptual experience will have to factor in at least some knowledge, belief, or awareness into the analysis of joint attention.

Acknowledgements Open Access funding provided by Projekt DEAL. We would like to thank the reviewers of *Synthese* for their very extensive comments. Bart Geurts was supported by the HSE University Basic Research Program.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Adamson, L. B., Bakeman, R., Suma, K., & Robins, D. L. (2019). An expanded view of joint attention: Skill, engagement, and language in typical development and autism. *Child Development*, *90*(1), e1–e18.

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, *55*(4), 1278–89.

Bruner, J. S. (1974). From communication to language: A psychological perspective. *Cognition*, *3*(3), 255–287.

Burge, T. (2005). Disjunctivism and perceptual psychology. *Philosophical Topics*, *33*(1), 1–78.

Byrne, A., & Logue, H. (2008). Either/or. In A. Haddock & F. Macpherson (Eds.), *Disjunctivism: Perception, action, knowledge* (pp. 314–19). Oxford: Oxford University Press.

Calabi, C. (2008). Winks, sighs and smiles? Joint attention, common knowledge and ephemeral groups. In H. B. Schmid, K. Schulte-Ostermann, & N. Psarros (Eds.), *Concepts of sharedness: Essays on collective intentionality* (pp. 41–58). Frankfurt: De Gruyter.

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

Campbell, J. (2005). Joint attention and common knowledge. In N. M. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 287–297). Oxford: Oxford University Press.

Campbell, J. (2009). Consciousness and reference. In A. Beckermann, B. P. McLaughlin, & S. Walter (Eds.), *The Oxford handbook of philosophy of mind* (pp. 648–662). Oxford: 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. 415–30). Cambridge, MA: MIT Press.

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

Carpenter, M., & Liebal, K. (2011). Joint attention, communication, and knowing together in infancy. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 159–182). Cambridge, MA: MIT Press.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63*(4), i–174.

Crane, T. (2006). Is there a perceptual relation? In T. S. Gendler & J. Hawthorne (Eds.), *Perceptual experience* (pp. 126–146). Oxford: Oxford University Press.

De Sá Pereira, R. H. (2016). Combining the representational and the relational view. *Philosophical Studies*, *173*(12), 3255–3269.

Eilan, N. (2005). Joint attention, communication, and mind. In N. Elian, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 1–33). Oxford: Oxford University Press.

Eilan, N. (2015). *Joint attention and the second person (draft)*. https://warwick.ac.uk/fac/soc/philosophy/people/eilan/jaspup.pdf. Accessed October 9, 2018.

Gallagher, S. (2010). Joint attention, joint action, and participatory sense making. *Revue de Phénoménologie*, *18*, 111–124.

Geurts, B. (2019). Communication as commitment sharing: Speech acts, implicatures, common ground. *Theoretical Linguistics*, *45*, 1–30.

Hobson, P. (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: Oxford University Press.

Hobson, P., & Hobson, J. (2011). Joint attention or joint engagement? Insights from autism. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 115–136). Cambridge, MA: MIT Press.

León, F., Szanto, T., & Zahavi, D. (2019). Emotional sharing and the extended mind. *Synthese*, *196*(12), 4847–4867.

Lewis, D. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Martin, M. G. F. (1997). The reality of appearances. In M. Sainsbury (Ed.), *Thought and ontology* (pp. 81–106). Milan: Franco Angeli.

Martin, M. G. F. (2004). The limits of self-awareness. *Philosophical Studies*, *120*(1–3), 37–89.

Moll, H., & Meltzoff, A. N. (2011). Joint attention as the fundamental basis of understanding perspectives. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 393–414). Cambridge, MA: MIT Press.

Moll, H., & Tomasello, M. (2007). Cooperation and human cognition: The Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 639–648.

Moore, C., & Dunham, P. J. (1995). Current themes in research of joint attention. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 15–28). Hillsdale, NJ: Lawrence Erlbaum.

Mundy, P. C. (2016). *Autism and joint attention: Development, neuroscience, and clinical fundamentals*. New York: Guilford Publications.

Nanay, B. (2014). Empirical problems with anti-representationalism. In B. Brogaard (Ed.), *Does perception have content?* (pp. 39–50). Oxford: Oxford University Press.

Nanay, B. (2015). The representationalism versus relationalism debate: Explanatory contextualism about perception. *European Journal of Philosophy*, *23*(2), 321–336.
Peacocke, C. (2005). Joint attention: Its nature, reflexivity, and relation to common knowledge. In N. M. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 298–324). Oxford: Oxford University Press.
Scaife, M., & Bruner, J. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*, 265–266.
Schellenberg, S. (2010). The particularity and phenomenology of perceptual experience. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *149*(1), 19–48.
Schiffer, S. R. (1972). *Meaning*. Oxford: Clarendon Press.
Seemann, A. (2011). *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience*. Cambridge, MA: MIT Press.
Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Lawrence Erlbaum.
Tomasello, M. (2014). *A natural history of human thinking*. Cambridge, MA: Harvard University Press.
Travis, C. (2004). The silence of the senses. *Mind*, *113*(449), 57–94.

Chapter 3

Paper II: Opening up the Openness of Joint Attention

Abstract

The ability to engage in joint attention, in which two individuals attend to the same object or event together, is considered fundamental for language learning, for understanding others and for joint actions. Joint attention is often defined as a mutually open, or transparent relation between co-attenders. But how should this openness be characterised? Two broad theoretical views have been proposed. One view reductively accounts for the mutual awareness characteristic of joint attention in terms of individual mental states and properties. According to non-reductive views, in contrast, mutual awareness is based on some primitive intersubjective relation, which is irreducible to the individual states of its relata. I argue that tensions in these approaches arise from the attempt to address both normative and cognitive explananda simultaneously. Both approaches are primarily designed to tackle the normative epistemological concerns of joint attention, and their problems arise when they conflate these concerns with psychological ones. Drawing from evidence in developmental and cognitive psychology, I outline the case for a cognitive-first account of joint attention based on a weaker notion of openness and mutual awareness. I conclude by assessing the epistemic implications of this account.

> "I think (and I think Hume did too) that, insofar as it's about the analysis of justification and the like, epistemology hasn't really got much to do with psychology." (Fodor, 2003, 4)

1 Introduction

Two people sit at a table with a piece of cake between them. They look at the cake, exchange glances and smile. They are thus both attending to the cake. Importantly, they are both at the same time aware of each other's attention. If one of them was to say "Grandma made it", the referent would be clear to both of them. This is a paradigmatic case of joint attention.

It is widely agreed that joint attention, in which two individuals attend to the same object or event together, plays an important role in language development and communication, in joint action, and in the progressive understanding that others can have different perspectives than our own. More generally, joint attention supports the development of mentalising, the ability to comprehend other people's mental lives (Moore & Dunham, 1995; Carpenter et al., 1998; Mundy, 2018). Within philosophy, joint attention has been considered essential for the ability to distinguish self from other (Bermúdez, 1998), the constitution of a common ground for communication (Clark, 1992), and the concept a shared objective world, where mind-independent objects are attended in common (Davidson, 1999; Campbell, 2011). However, there is scant agreement on how joint attention should be analysed and how it is involved in all these capacities.

This paper aims to contribute to theoretical research on joint attention by examining the notion of a shared "openness" between co-attenders. Most researchers agree that a key feature of the state of joint attention is that it is public or "out in the open" among the co-attenders. It is fully and immediately "transparent", or "mutually manifest", to them that they are jointly attending to the same object or state of affairs (Bakeman & Adamson, 1984; Tomasello, 1995; Eilan, 2005; Peacocke, 2005; Calabi, 2008; Campbell, 2011; Carpenter & Liebal, 2011). This feature puts the jointness in joint attention and distinguishes it from cases where we attend to the same object unilaterally, completely unaware of each other. Co-attenders are *mutually aware* of their shared attention to the same target. But how should this notion of mutual awareness be analysed? Two broad theoretical views have been proposed. One view reductively accounts for the mutual awareness characteristic of joint attention in terms of individual mental states and properties. According to

non-reductive views, in contrast, mutual awareness is based on some primitive intersubjective relation, which is irreducible to the individual states of its relata.

I have two goals in this paper. First, I argue that much of the debate between reductive and non-reductive views arise through the conflation of two distinct explanatory aims: the aim of explaining the normative role of joint attention in justifying joint endeavours, and the cognitive aim of explicating the psychological capacities involved in joint attention. Reductive and non-reductive approaches are primarily concerned with the normative aim, and their problems arise when they extend their scope to tackle the cognitive aim (section 2). Both approaches, I argue, turn out to be conceptually equivalent when the focus is strictly normative. Second, I outline the case for a cognitive-first account of joint attention based on a weaker notion of mutual awareness, in line with empirical evidence from developmental psychology (section 3). I suggest that mutual awareness in joint attention is not something co-attenders must arrive at, but that it is often implicitly assumed, and that we must (un)learn that other people may not attend to the same things we attend. I conclude by returning to the normative question, assessing how this proposal can address the epistemic justificatory role of joint attention in our social lives.

2 Assessing the reductive and non-reductive views

2.1 Explanatory aims

It is commonly assumed that a primary functional role for joint attention is to provide a rational basis for further coordinated activities between agents. Joint attention is thus considered a necessary but not sufficient condition for individuals to engage in collective activities (2003); and plays an essential role in the constitution of the common ground involved in communication (Clark, 1992; Bruner, 1995; Tomasello, 2008). John Campbell shares in the current consensus by associating the openness of joint attention with this functional role:

> [W]hatever else is true of it, joint attention has an "openness" about it — there's some sense in which the situation is "open" to both attendees in a case of joint attention — in virtue of which joint attention ordinarily plays a distinctive role in rational, coordinated action. (Campbell, 2011, 417)

On the other hand, joint attention is a psychological phenomenon, and any theoretical analysis will be constrained by psychological plausibility. For example, Christopher Mole (2017) notes that, "since joint attention is achieved by young children, its achievement cannot plausibly be thought to make any sophisticated intellectual demands" (cf. Roessler, 2005). We need to spell out what are the mental and cognitive requirements for joint attention: how is openness achieved? An account of the openness of joint attention is therefore seen as a necessary step in addressing these two different explananda:

i. Normative: How does joint attention rationally support joint beliefs and joint actions, and yields shared knowledge of our environment?

ii. Cognitive: What cognitive capacities and mental processes or understanding are involved in joint attention, which even infants and young children may be capable of having?

Though they needn't be, these two questions are often tackled at once in the current debate. Supposedly, the openness of joint attention must be conceptualised in such a way that it provides a normative epistemic basis for rational behaviour regarding the co-attended object, and simultaneously shed light on the mental wherewithal necessary to achieve said openness. I examine the contrast between reductive and non-reductivist accounts of joint attention in light of these two different explanatory aims. Both accounts take the normative question (i) as a starting point and motivation. It is usually assumed that the functional role of joint attention is to provide the categorical grounds that allow subjects to engage rationally in joint projects concerning the jointly attended object, including communicative speech referring to the object. In the rest of this section, I show how the problems for both accounts arise only when they are taken to simultaneously address the cognitive question (ii).

2.2 The reductive approach

The two broad theoretical approaches to joint attention tend to give priority to slightly distinct notions of openness. On an epistemic interpretation of openness, a situation S is open between A and B when S is epistemically shared between A and B. The reductive approach is paradigmatic in taking epistemic openness as the starting point. While this view does not deny the experiential character of joint attention, it holds instead that this character is better accounted for in terms of its epistemic structure. According to this approach, epistemic openness can be explained by reducing it to the individual mental states of each co-attender. Reductive views tend to analyse the openness of

joint attention in terms of common knowledge or similar notions such as mutual awareness, constructed in line with the analyses from Lewis (1969) and Schiffer (1972). On this definition, A and B are jointly attending to x when they are mutually aware, or enjoy common knowledge, that each of them is attending to x. Common knowledge is defined as giving rise to iterative structures like the following:

p is common knowledge between A and B iff

- A knows that *p*,
- B knows that *p*,
- A knows that B knows that *p*,
- B knows that A knows that *p*,
- and so on, *ad infinitum*.

On this construal, joint attention involves nested psychological states that both A and B would have to entertain about each other. This is the hallmark of recursive mindreading, where one subject attributes mental states to another, which, in turn, refer to the first subject's own mental states. Typical criticisms of this approach centre on the requirement of recursive mindreading. One argument from phenomenology notes that the openness in joint attention is immediate and effortless, so that it is implausible that joint attention requires recursive mindreading (e.g. Gallagher, 2011). Related arguments point to the computational complexity of recursive mindreading, arguing that it is intellectually demanding and psychologically implausible even for adults (Eilan, 2005; Campbell, 2018).

One key argument against the reductive approach based on recursive iterations appeals to "coordinated-attack" scenarios (Wilby, 2010; Campbell, 2005, 2018).[1]

> *Coordinated-attack scenario*: Two individuals in separate booths must both attack the same target among many, at the same time. For this, they will have to coordinate their individual actions. Additionally, there is always a non-zero chance of distorted communication between booths, so that when one individual has chosen a target and communicates this to the other, they will not know

[1]The coordinated-attack problem was first introduced in its current form by Akkoyunlu et al. (1975) in the context of dynamical systems engineering and, while it soon became a key fixture in epistemic logic, it had little to do with mental representations and psychological processes.

> for sure whether the communication has been received. Supposing that the second individual does receive the message specifying the target, they could, in turn, send a message back to confirm receipt of the target. But, again, they will not be sure whether this confirmation has been received. To make a coordinated attack rational, both individuals will have to go through an infinite iteration of messages confirming the preceding confirmation.

Coordinated-attack scenarios are often assumed to show that no finite iteration of inferences will allow the participants to engage in rational coordinated attack and secure victory. Yet in normal situations, where we are both present in the same physical space, we can easily arrive at a successful coordinated outcome. When both individuals and the target are all co-present, as in most normal circumstances, then "everything is out in the open to such an extent that we can rationally attack" (Campbell, 2005: 292). So how do we do it, given that we cannot perform infinite inferences?

For Wilby (2010), the coordinated-attack problem "highlights that once one gets embroiled in *supposing* that an act of transparent communication or shared knowledge requires a set of hierarchical to-ing and fro-ing about who knows what, then there will be no end to the matter" (my emphasis). Going up only two or three levels up the recursive chain will not suffice. As Wilby (2010) and Campbell (2018) note, for any level in the chain, that level does not lead to openness or mutual awareness, or else will require a further step in the recurse chain[2]. The conclusion of the coordinated-attack argument is what Wilby calls as disastrous "paradox" for the reductive approach: mutual awareness requires an infinite recursion of overlapping mental states, but this requirement is psychologically implausible.

There are two points where the argument from coordinated-attack fails. First, this rendering of the "paradox" relies on the assumption that on the iterative approach, openness itself is the *result* or *end-point* of the iterations. On this assumption, it follows that the openness in normal joint attention scenarios is similar to the openness which the individuals in our separate booths are infinitely pursuing, which spells problems for the reductionist. Yet the reductionist is not committed to make such assumption in the first place. Second, there is nothing in the coordinated-attack scenario that implies that

[2]One may suggest that people just reason two or three levels up the hierarchy and then conclude their knowledge is shared. This is an empirical, not a logical, suggestion; and one with no evidential support (e.g., Liddle & Nettle, 2006; Thomas et al., 2014). The approach implicit in this suggestion is normative-first: start with a logical, normative theory, and truncate it to fit what (we think) humans can do. I suggest adopting instead a cognitive-first approach and evaluate any normative implications thereafter.

the iterations must be actually *represented* in the mind of each individual in a case of joint attention. This construal of the argument relies on the convergence of both normative and cognitive aims described above. The view that a reductive approach to joint attention involves recursive mindreading is based on normative analyses of common knowledge such as those by Lewis and Schiffer. Traditionally, their approach presumes that there are situations with some finite conditions, out of which the infinite iterations logically follow. What makes a situation a common knowledge situation are those finite conditions, not the recursive iterations themselves, nor their result or end-point. In other words, the iterations arise as logical implications which follow from some given finite situation, and which are not necessarily represented in anyone's reasoning (Lewis, 1969). Coordinated-attack scenarios only show that the individuals in their separate booths lack the appropriate finite conditions.

To date, providing a good account of those finite conditions has been proved somewhat problematic, but only when, in addition, such conditions must also account for the psychological processes and mental wherewithal necessary for joint attention or common knowledge. If we do away with this cognitive explananda, we are left only with the normative question. Under the normative aim, the openness of joint attention can be seen as a purely normative epistemological notion, and there is nothing psychological about it. Analyses of joint attention and mutual awareness in the style of Schiffer's and Lewis' are not necessarily committed to an infinite regression of mental states. They are not in principle committed to any view about psychological processes at all, and so assuming that they must involve the performance of an infinite chain of mental states is a misapplication of a normative analysis into a psychological straightjacket.[3]

This conclusion leaves unanswered what mutual awareness is, psychologically speaking, as a mental state enjoyed by both adults and infants. In other words, the upshot is that we are giving up the cognitive explanandum of joint attention.

2.3 The non-reductive approach

Prominent non-reductive approaches concentrate instead on experiential openness (Campbell, 2005, 2018; Seemann, 2019; Wilby, 2010). On a non-reductive view, a situation S is open between A and B when S is fully present

[3] For a purely normative treatment of phenomena that share the openness under consideration, particularly the common ground between interactants and its role in communication, see Geurts (2019).

to the consciousness of A and B (see Calabi, 2008). It is in virtue of its phenomenal character that joint attention plays an epistemic role in justifying shared beliefs and joint activities. On this approach, there is a primitive intersubjective relation behind joint attention, which cannot be analysed any further. John Campbell thus proposes that joint attention is a primitive type of conscious state (2005; 2018). Just as the object you see can be a constituent of your experience, so too it can be a constituent of your experience that the other person is, with you, jointly attending to the object. Naomi Eilan argues that joint attention is grounded in experiences of "you-awareness" and "communication-as-connection", which are primitive conscious states (2015). Following Campbell, Axel Seemann (2019) argues that our perceptual experience during joint attention is a primitive joint state. What each of us experiences cannot be reduced to our individual psychological states, but is determined by the triadic spatial arrangement between us and the common object of our attention (Seemann, 2019, 75).

Campbell's analysis, in particular, is based on the premise that joint attention can be explained fully in terms of perceptual experience, and thus is not susceptible to an explanation in terms of the knowledge, beliefs, or awareness of the two participants (Campbell, 2018, 120). For this reason, it has proved to be an attractive theoretical position for cognitive and developmental psychologists (e.g. Moll & Meltzoff, 2011; Hobson & Hobson, 2011). One serious criticism of this approach is that it simply embeds in the analysis of openness or mutual awareness the property that is to be explained, i.e., the openness of joint attention (Peacocke, 2005). Further problems arise due to the narrow focus on perceptual experience. Since the non-reductive approach allows for dissociations between one's perceptual experience and one's beliefs about that experience, it also allows for the possibility that one is engaged in joint attention but mistakenly believes that is not (Battich & Geurts, 2020).

As it is perhaps already evident, the key motivation for non-reductive approaches is to address the normative explananda of joint attention, and its problems — or rather, its limitations — arise when it is also taken to provide insight into the cognitive question. When focusing exclusively on the normative question, however, the openness of joint attention becomes merely a description of phenomenal aspects of experience. Being primitive, these aspects cannot be further explained. But neither do they inform us about the psychological capacities behind joint attention. On the non-reductive approach, joint attention is treated from a third person point of view, as it merely asks whether an external *ascription* that A and B jointly attend to x is true. A description of a primitive intersubjective relation between co-attenders can make such ascriptions true, but this description will be entirely silent on the psycholo-

gical states of each individual (cf. Schmitz, 2015, 239). The upshot, of course, is that we are giving up the cognitive explanandum (aim ii).

2.4 Conceptual equivalence between iterative and primitivist approaches

For primitivists like Campbell (2005) and Seemann (2019), the state of mutual awareness in joint attention is a factive state. You cannot be aware that of the other person is currently co-attending with you to the same perceptual target, unless that person is, in fact, a co-attender. Of course, you could be wrong about your co-attendance, but then you would not be in a state of joint attention. The factive character of the mental states of each co-attender is also usually assumed for the reductive, iterative approach. Two people are mutually aware of *p* only when both of them are equally justified to follow the infinite logical implications of their joint epistemic state. Interestingly, the assumption that the mental states of each co-attender are factive implies that the iterative and primitivist views are conceptually equivalent — at least under some versions of each view. In particular, the equivalence holds for Schiffer's analysis of common knowledge, commonly taken as a paradigm of the reductive, iterative approach when applied to the openness in joint attention. The brilliance of Schiffer analysis is that it proposes a *finite basis* for common knowledge, out of which the iterations would follow logically. Therefore, it does not necessitate an infinite recursion of mental states, to be represented in the minds of each individual. On Schiffer's analysis, A and B mutually know that *p* iff there are properties F and G such that:

1. A is F.
2. B is G.
3. Both being F and being G are sufficient for knowing that *p*, that A is F, and that B is G.
4. For any proposition *q*, if both being F and being G are sufficient for knowing that *q*, then both being F and being G are sufficient for knowing that both being F and being G are sufficient for knowing that *q*. (Schiffer, 1972, 34-5)

Given this finite base, the infinite number of iterations characteristic of common knowledge can be generated by feeding (3) to the recursive clause in (4), and reapplying (4) to each new result over and over. This analysis relies heavily on the generating properties F and G, which, Schiffer proposes, refer to the property of being "a visibly 'normal', open-eyed, conscious person":

> If a "normal" person (i.e., a person with normal sense faculties, intelligence, and experience) has his eyes open and his head facing an object of a certain size (etc.), then that person will see that an object of a certain sort is before him. (Schiffer, 1972, 31)

Moreover, people know that normal people will behave in this way, and they can easily tell when someone is normal (Schiffer, 1972, 33). For this analysis to work, the property of "being normal", however, must be relativized to the specific situation in which the co-attenders currently are (Wilby, 2010). Some situations will require, for example, that assumptions about normal hearing, rather than normal sight, be included in the normality properties F and G to allow for common knowledge towards an auditory event in the environment. Importantly, Wilby (2010) has shown that F and G are more intimately related than Schiffer initially presumed. Schiffer's four clauses, in particular (3), together with the facticity assumption (which makes knowledge a factive state by definition: if X knows *p*, then *p* is true), imply that the two generating properties F and G are necessary and sufficient conditions for each other:

1. Both being F and being G are sufficient for knowing that *p*, that A is F, and that B is G (assumption from Schiffer)
2. If X knows that *q*, then *q* is true (facticity assumption)
3. If A is F, then A knows that B is G (from 1)
4. If B is G, then B knows that A is F (from 1)
5. If A is F, then B is G (from 2 and 3)
6. If B is G, then A is F (from 2 and 4)
7. A is F iff B is G (from 5 and 6) (Wilby, 2010, 91)

Given the biconditional relation between F and G, these properties can be logically replaced by a primitive relational property H, so that H iff F and G. Thus, Schiffer's analysis of common knowledge turns out to be logically equivalent to an analysis including the single intersubjective property H, where A and B are both H, and each of them can only be H when the other person is likewise H. But now the analysis includes a primitive intersubjective element irreducible to the mental states and properties of each individual. Schiffer's analysis of common knowledge is conceptually identical to a primitivist analysis. Common knowledge, under Schiffer's analysis together with the facticity assumption, is a relational state irreducible to the individual cognitive states of the individuals in question (Wilby, 2010, 92). It is important to be clear about the implications of this equivalence. Wilby suggests that the equivalence, together with the psychological implausibility of the iterative approach, arbitrate in favour of the primitivist approach to common knowledge and related

epistemic notions, such as mutual awareness and joint attention. This suggestion is unwarranted, however. Wilby's argument demonstrating the equivalence between the two views constitutes a redrawing of the normative aspects of common knowledge. In principle, the primitive relational property H has little to do with psychology and the actual mental states of an individual (Wilby, 2010, 93, admits as much).

One could go further, of course, and interpret H as an irreducible joint *psychological* state. But this interpretation only brings back the limitations of the primitivist approach: the openness of joint attention becomes an irreducible psychological state, and, as Wilby himself notes, it is suspect how much explanatory work such irreducible notion can play in psychological theories and experimental research. In particular, the cognitive explananda is left untouched: what mental processes and understanding are involved in joint attention, which even infants and young children may be capable of having? This defeatist outcome is due, in part, to the equivalence relying on the facticity assumption. While this clause is commonly assumed for knowledge (Williamson, 2000), it arguably does not hold for psychologically-determined states such as beliefs and awareness. It is dubious that a person's *psychological* sense of being in a situation of joint attention leads, as a matter of logical necessity, to joint attention being true.

The conceptual equivalence of iterative and primitivist analyses holds when both analyses are strictly considered as epistemic normative theories. Their unreconcilable differences arise only when they are taken to address, in addition, the cognitive question regarding the psychological processes behind joint attention. If we are interested in the psychological states of real humans and children, and not just in the rational states of epistemic agents, then we should acknowledge the limitations of such "normative-first" analyses of joint attention. Neither of the two approaches makes it possible to address the cognitive instead of the normative question.

3 A cognitive-first approach to joint attention

3.1 Mutual awareness is assumed

In 2003 Michael Tomasello remarked that child language acquisition is not a logical problem, but an empirical one. He urged that a theory of human linguistic competence should be based less be on analogies to formal languages, and more on empirical research in the cognitive sciences (2003, 328). Unfortunately, Tomasello himself didn't fully apply this dictum to the topic of common knowledge and joint attention, assuming, along with many others, that it

must involve either something akin to recursive mindreading, or some primitive non-analyzable plural "we" subject, which is in turn presumably produced by as yet undiscovered unconscious mechanisms (cf. Zawidzki, 2013):

> From early on as well, infants communicate with others referentially, inviting them to jointly attend to something, and this requires recursive inferences about mental states embedded in mental states. (Tomasello, 2019, 44)

Given the narrow "normative-first" focus of traditional approaches to joint attention, however, I propose that, if we are interested in the cognitive question, there are no strong reasons to assume a priori that the key feature of joint attention is a fully symmetric epistemic or experiential openness. Instead, I propose that a more fruitful approach to describe the triadic interaction of joint attention is to concentrate on what factors or aggregate of factors each individual co-attender is responding to, so that this interaction can be established, without presuming in advance the nature of the epistemic or phenomenal sophistication they must achieve. In this section, I provide the outline of a "cognitive-first" approach to assess the jointness of joint attention. The starting point is to leave normative concerns on the side for the time being. The aim of a cognitive-first approach is not to arrive at a *justification* for the mental states of an individual during joint attention. We are not (yet) concerned with the rationality of those mental states.

In a nutshell, I suggest that mutual awareness in joint attention is neither a primitive nor reductive intersubjective relation that co-attenders must arrive at, but that it is often implicitly assumed, and that we must (un)learn that other people may not attend to the same things we attend, or may not share the same perceptual knowledge we are currently enjoying. This view is supported by research in developmental and cognitive psychology. Two-year-old children typically assume that an adult interacting with them will share their perceptual perspectives (Moll & Meltzoff, 2011; Epley et al., 2004). In one particular experiment, two-year-old children shared visual attention of two objects with an adult, one by one. The child was then presented with a third object, which the adult could not see. In one condition, this was because the adult was present behind a barrier, so that they didn't have visual access to the third object but continued to communicate verbally while the child inspected the object. In another condition, the adult was completely absent from the room when the third object was shown to the child. The task was to identify which of the three objects was new for the adult when she explicitly requested to the child for the "one she has not seen before". Children were able to correctly select the new object when the adult had been absent from the room,

but they were not able to differentiate between new and old objects when the adult was behind the visual barrier and could not see the object but still engage with them verbally (Moll et al., 2011).

Moll and colleagues interpret these results as showing that physical co-presence and some form of minimal engagement is enough for children to assume that they are sharing their perceptual experiences with the adult (see also Hobson & Hobson, 2011). These findings are in line with, for example, the everyday experience of a child talking in the telephone and assuming that the other person is aware of what they are pointing to. The impulse to assume openness or sharing experiences or knowledge is not restricted to children alone. A similar phenomenon is observed in adults, where someone's own knowledge will affect, however implicitly, their ability to reason about another person's beliefs (Epley et al., 2004). Referred to as the "curse of knowledge", people are egocentrically biased to assume that others know what they themselves know (Birch & Bloom, 2007; Farrar & Ostojić, 2018).

On the view I propose, then, during joint attention co-attenders merely have to assume, pre-reflectively, that their attention to the same object is shared with someone else's. Contrary to the traditional reductive view, attaining perceptual common knowledge towards the same object is not cognitively taxing, but curbing it down is: taking into consideration whether other people do not share your object of attention is cognitively demanding, at least during development and in novel situations with no prior precedents. Unlike non-reductive views which posit some intersubjective primitive phenomenon behind joint attention, the view I propose is anchored on the individual. It concerns the mental processes that an individual A must go through in order to say that she is jointly attending with B to x (and is mutually aware with B about so being in joint attention to x). Of course, B and her mental processes will often come into the picture too. However, in this analysis, B musn't necessarily be a minded individual. A can engage in joint attention with, e.g., a computer avatar or with animals. Whatever B knows or is aware of is not constitutive of this analysis. This account concerns the mental and psychological states of A alone, so that we can say that A takes herself, from her perspective and her practical purposes (though not necessarily consciously) to be jointly attending with B to x.

This proposal allows us to get a grip on the cognitive question without as yet being misled by normative concerns. What cognitive capacities and mental processes or understanding are involved in joint attention? What minimal capacities are necessary to pre-reflectively assume that one is sharing attention to the same object with others? Based on the studies by Moll and col-

leagues, we can start with a set of minimal cues that I as a co-attender should be capable of recognising:

- You are a separate individual from me.
- You are physically present.
- You engage with the world as I do.

Based on these cues, I can form the subpersonal representation that the attention towards x is shared. This cognitive process could be paraphrased as "I have a certain relation to x, and since you and I are so similar, (I assume) you have it too." It becomes clear that the notion of mutual awareness I am using here is considerably weaker than the notion used by reductive and non-reductive approaches. When I assume mutual awareness, this does not imply that this assumption must be fully conscious, reflective or deliberate, or that I have to consciously entertain the proposition that you are similar to me. It does not even require having a concept of mutual awareness. The only requirement is that I recognise and respond to the cues that you provide by implicitly assuming that you engage with the object in the same way I engage to it. Such recognition and response is plausibly supported by subpersonal sensory-motor and affective processes (cf. Reddy, 2010). Conceptual and reflective awareness of this engagement plays no role.[4] Of course, I could become reflectively aware of our joint engagement towards x. Usually, this occurs when the assumption of joint attention breaks or misfires. If I say to you "Grandma made it", and you show no comprehension of the intended referent, I can become retroactively conscious that I assumed, incorrectly, that we were looking at the cake together. When interaction fails, an individual may learn to tone down their assumptions of openness in similar future interactive situations, and revaluate the set of cues that trigger their assumption of mutual awareness in those situations.

Uncovering the set of cognitive capacities behind joint attention is an at root an empirical project, not a purely conceptual one. But conceptual clarity regarding different explanatory aims can assist with this empirical project. For this reason, while the present proposal is yet underdeveloped as a full response to the cognitive question and cannot be the complete story, it serves as an illustration of a cognitive-first approach to joint attention and its openness.

[4] In contrast to both Campbell (2005) and (2005), I do not aim to characterise the phenomenal experience of openness itself. I remain noncommittal to what the phenomenology of this assumption might be. However, it is important to note that the functional role of assuming mutual awareness, in the weaker sense used here, can be carried out without invoking its phenomenological aspect.

3.2 Epistemic implications

Mutual awareness in joint attention is implicitly assumed. How would this approach fare in accounting for the epistemic justificatory role that joint attention is taken to play in human social lives? The normative aim is to explain how joint attention supports joint beliefs and joint actions, and yields shared knowledge of our environment. The proposed account, I suggest, shifts the epistemic normative question from explaining how individuals can attain a relation of mutual awareness in joint attention that justifies shared knowledge of the perceptual environment, to explaining how individuals abstain from defaulting to possibly erroneous assumptions of mutual awareness. This approach concerns the mental and psychological states of each individual alone, so that all we can say, regarding A, is that A attends to x, and that A takes herself to be attending to x together with B. Whether A is actually justified to take herself to be in a situation of joint attention is a further question, which will depend on factors *external* to her psychology. More precisely, it will depend on what is happening inside the mind of B: A is justified to take herself to be in a situation of joint attention with B iff B attends to x, and B takes herself to be attending to x together with A. Conversely, B is justified to take herself to be in a situation of joint attention with A iff A attends to x, and A takes herself to be attending to x together with B. We can now spell out a normative epistemic account of joint attention.

A and B jointly attend to x iff

1. A attends to x.
2. B attends to x.
3. A takes herself to be attending to x together with B.
4. B takes herself to be attending to x together with A.

This account seems at first circular. A critic may point out that conditions (3) and (4) already presuppose what we are trying to analyse, i.e., the jointness of joint attention. On the other hand, a primitivist proponent might in turn retort that (3) and (4) should be taken as primitive conditions, which cannot be further analysed. I disagree with both views. It is important here to recall the distinction between cognitive and normative aims and explananda. One thing is to have a psychological state of being in joint attention. A different thing is to outline the justifications for that state. To aid maintain this distinction, I suggest differentiating between *psychological* joint attention and *normative* joint attention. The normative notion of joint attention consists of conditions

(1) - (4). Conditions (3) and (4) themselves, however, refer strictly to the psychological states of A and B, states which need not include any concept, reflection or awareness as to the normative force they play when all conditions (1) - (4) are realised. In other words, the aetiology of these states need not involve the normative notion of joint attention. There is, therefore, no circularity, and there are no primitive unanalysable intersubjective states. Conditions (3) and (4), as psychological states, can and should be further analysed. The empirically-based view proposed in the previous section is an attempt, after all, to sketch the cognitive capacities that go into the psychological state of pre-reflectively assuming that one is sharing attention to the same object with someone else.

Returning to the normative question, how are condition (1) - (4) realised? As noted above, the normative problem for A (as for B, mutatis mutandis) is not to *arrive* at the state that she is jointly attending to x with B. The problem A faces is to *avoid* defaulting to her prior assumption that she jointly attends to x with B, when that default should not be made — that is, in cases where B does not share her attention. The problem A faces is to avoid defaulting to her prior assumption that (3), when (3) should be rejected. There are two basic ways in which A's having a psychological state of assuming shared attention to x with B is not rationally justified:

i. Both conditions (2) and (4) don't hold.

ii. Condition (2) holds but (4) doesn't. B attends to x, but does not have a psychological state of being attending to x together with A.[5]

Since A's psychological state of sharing attention to x with B does not need to include conditions (2) and (4), its rational justification is external to A's psychology. Her psychological state can be merely based on the following implicit reasoning: "I have a certain relation to x, and since you and I are so similar, (I assume) you have it too." Nothing requires A to reason any further. If she stops here, she takes herself to jointly attend to x with B (as a matter of *psychology*, that is). She may, of course, be epistemically wrong and unjustified. This is one possible stage of epistemic failure. Given (i) and (ii), to override defaulting to assumptions of mutual awareness, however, A will have to estimate the probability that B does not attend x, and the probability that B's attention is not being shared (see Siposova & Carpenter, 2019).

[5]What about a situation where condition (4) holds but (2) doesn't? Whether this situation can ever occur will depend on whether (4) necessarily entails (2), which, in turn, will depend on the particular theory of perceptual attention endorsed. For example, one could erroneously assume oneself to be attending to x but be, in fact, attending to y. For simplicity, however, in this paper I assume that the entailment holds necessarily, so that the situation where condition (4) holds but (2) fails cannot occur.

These estimates constitute two further psychological factors that modulate the rationality of her assumption of mutual awareness (or her withholding the assumption). On the proposed hypothesis, estimating these probabilities is cognitively demanding — though such process can of course (and likely it does) occur unconsciously and pre-reflectively. Depending on her prior experience in domains involving objects such as x and people such as B, A may be more or less sensitive to relevant information for estimating these probabilities. According to her estimates, then, A will reject her assumption of mutual awareness and conclude that there is no joint attention with B, or A will keep her assumption. The accuracy of her estimates, however, is a matter of degree, and cannot put A in a fully justified and rational state of *normative* joint attention, since these estimates will never fully encompass, from the psychological stance of A alone, the truths of conditions (2) and (4). These two conditions are external to A, and therefore beyond the ken of a strictly psychological standpoint. Paraphrasing Herbert Clark (1996, 96), a fully rational state of *normative* joint attention, consisting of conditions (1) – (4) above, can only be held by an omniscient being. The rationality of A's assumption of mutual awareness is, on this account, a matter of degree.

Does this mean that joint attention's epistemic justificatory role in our social lives is in jeopardy? Although the answer will ultimately depend on how strict we make the notion of epistemic rationality, I do not see compelling reasons to conclude that it doesn't. There is no a priori need to assume that the functional role of joint attention is to provide logically irrefutable grounds allowing individuals to engage rationally in joint endeavours concerning the jointly attended object. This assumption not only ignores psychological plausibility, but it may also leave aside aspects of strategic rationality at play in joint attentional scenarios (cf. Todd & Gigerenzer, 2012). Williams James already noted that "the logic of belief and knowledge" is too abstract to tackle many epistemic situations. We cannot simultaneously attempt to believe as many truths as possible and as few falsehood as possible. According to James, we need to implicitly or explicitly weight the value of avoiding false positives against false negatives (James, 1956; cf. Van Fraassen, 2002, 88).

In a football game, for example, if I assume joint attention with an opponent toward the ball and get it wrong, I may not incur any significant costs, given my practical interests (and I may never realise I got it wrong). But if I don't assume joint attention, and should have (i.e. conditions (1), (2) and (4) hold, but A still fails to assume joint attention), the consequences could be drastic. Here, a false negative is more pernicious than a false positive. The coordinated attack scenario is an extreme case in the opposite direction: if I assume joint attention and get it wrong, then we both stand to incur high

costs. But if I don't assume joint attention, and should have, the costs are negligible. Here, a false positive is more pernicious than a false negative. It would seem rational for an agent to be sensitive to the pay-off structure of a particular situation. A notion of epistemic justification in joint attention based of strict objective accuracy, consisting of conditions (1) - (4), cannot capture such strategic aspects. A psychological state of being attending to an object or event together with another person may thus serve a functional role in action, without that state itself being necessarily accurate or irrefutably justified.

4 Conclusion

People effortlessly engage with others in activities that require attending together to some object or event. This ability of coordinated joint attention is considered to be fundamental to many aspects of human development, cognition, and interaction. It is widely held that joint attention is essentially public, or "out in the open". Going beyond the metaphor of openness, however, requires a proper account of the mutual awareness that underlies joint attention. Current accounts, I have argued, fail to distinguish between two distinct explanatory aims when theorising on the openness of joint attention. One aim is normative: how should the openness of joint attention be characterized to account for its epistemic significance? Engagement in joint attention provides a rational basis for coordinated actions and shared knowledge about the world and others. A distinct explanatory aim is cognitive: what cognitive capacities and mental processes or understanding are involved in joint attention? Both reductive and non-reductive accounts of the openness in joint attention are primarily concerned with the normative aim, and their tensions arise when they extend their scope to address the cognitive aim.

Drawing from empirical research in infants and adults, I suggest that the openness in joint attention is not something that co-attenders must arrive at, but is implicitly assumed. On this hypothesis, given the right sort of cues, people will tend to assume, often without any conscious reflection, that they are attending to some object or event together with someone else. Being able to entertain this assumption in the first place will require a set of minimal cognitive capacities, including the understanding that the other individual is a separate, live organism with their own goals, and that they engage with the world in a similar way to oneself. Ultimately, however, uncovering the set of cognitive capacities behind joint attention is an empirical project, not a purely conceptual one. Distinguishing between cognitive- and normative-first approaches to joint attention allow us to distinguish between two different no-

tions of the concept. Joint attention as a *psychological* state requires an analysis at the level of the mental and psychological processing of the individual. Joint attention as a *normative* state — at the level of a "space of reasons", and pertaining to justified rational epistemic states of the individuals — will have to include externalist conditions outside a person's psychology. On this proposal, an individual cannot ever be fully aware of having rationally justified joint attention toward a common object with a third party, nor can they ever be fully aware of all the factors that make their state of joint attention rationally justified (cf. Sperber & Wilson, 1995, 19-20). They can, at best, form more or less accurate estimates of these factors. These estimates may be sufficient to account for the functional role of joint attention in supporting social behaviours and joint actions. If we are interested in providing a psychologically expedient construct of joint attention (Eilan, 2005; Wilby, 2010; Campbell, 2018), these considerations suggest that the notion of a fully normative state of openness in joint attention may well be cast aside.

5 References

Akkoyunlu, E. A., Ekanadham, K., Huber, R. V., Akkoyunlu, E. A., Ekanadham, K., & Huber, R. V. (1975). Some constraints and tradeoffs in the design of network communications. In *Proceedings of the Fifth ACM Symposium on Operating Systems Principles*, volume 9 (pp. 67–74). New York, NY: ACM Press.

Bakeman, R. & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, 55(4), 1278–89.

Battich, L. & Geurts, B. (2020). Joint attention and perceptual experience. *Synthese*. doi: 10.1007/s11229-020-02602-6.

Bermúdez, J. L. (1998). *The Paradox of Self-Consciousness*. Cambridge, MA: MIT Press.

Birch, S. A. & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, 18(5), 382–386.

Brinck, I. & Gärdenfors, P. (2003). Co-operation and communication in apes and humans. *Mind and Language*, 18(5), 484–501.

Bruner, J. S. (1995). From joint attention to the meeting of minds. In C. Moore & P. J. Dunham (Eds.), *Joint Attention: Its Origins and Role in Development* (pp. 1–14). Hillsdale, NJ: Lawrence Erlbaum.

Calabi, C. (2008). Winks, sighs and smiles? Joint attention, common knowledge and ephemeral groups. In H. B. Schmid, K. Schulte-Ostermann, & N. Psarros (Eds.), *Concepts of Sharedness: Essays on Collective Intentionality* (pp. 41–58). Frankfurt: De Gruyter.

Campbell, J. (2005). Joint attention and common knowledge. In N. M. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint Attention: Communication and Other minds. Issues in Philosophy and Psychology* (pp. 287–297). Oxford: 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. 415–30). Cambridge, MA: MIT Press.

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

Carpenter, M. & Liebal, K. (2011). Joint attention, communication, and knowing together in infancy. In A. Seemann (Ed.), *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience* (pp. 159–182). Cambridge, MA: MIT Press.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63(4), i–174.

Clark, H. H. (1992). *Arenas of Language Use*. Chicago, IL: The University of Chicago Press.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Davidson, D. (1999). The emergence of thought. *Erkenntnis*, 51(1), 511–521.

Eilan, N. (2005). Joint attention, communication, and mind. In N. Elian, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint Attention: Communication and Other Minds. Issues in Philosophy and Psychology* (pp. 1–33). Oxford: Oxford University Press.

Eilan, N. (2015). Joint Attention and the Second Person (draft). https://warwick.ac.uk/fac/soc/philosophy/people/eilan/jaspup.pdf.

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, 40(6), 760–768.

Farrar, B. G. & Ostojić, L. (2018). Does social distance modulate adults' egocentric biases when reasoning about false beliefs? *PLOS ONE*, 13(6), e0198616.

Fodor, J. A. (2003). *Hume Variations*. Oxford: Clarendon Press.

Gallagher, S. (2011). Interactive coordination in joint attention. In A. Seemann (Ed.), *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience* (pp. 293–305). Cambridge, MA: MIT Press.

Geurts, B. (2019). Communication as commitment sharing: Speech acts, implicatures, common ground. *Theoretical Linguistics*, 45(1-2), 1–30.

Hobson, P. & Hobson, J. (2011). Joint attention or joint engagement? Insights from autism. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 115–136). Cambridge, MA: MIT Press.

James, W. (1956). The will to believe. In *The Will to Believe and Human Immortality* (pp. 1–31). New York, NY: Dover Publications.

Lewis, D. (1969). *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Liddle, B. & Nettle, D. (2006). Higher-order theory of mind and social competence in school-age children. *Journal of Cultural and Evolutionary Psychology*, 4(3), 231–244.

Mole, C. (2017). Attention. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Stanford, CA: Metaphysics Research Lab, Stanford University.

Moll, H., Carpenter, M., & Tomasello, M. (2011). Social engagement leads 2-year-olds to overestimate others' knowledge. *Infancy*, 16(3), 248–265.

Moll, H. & Meltzoff, A. N. (2011). Joint attention as the fundamental basis of understanding perspectives. In A. Seemann (Ed.), *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience* (pp. 393–414). Cambridge, MA.: MIT Press.

Moore, C. & Dunham, P. J. (1995). Current themes in research of joint attention. In C. Moore & P. J. Dunham (Eds.), *Joint Attention: Its Origins and Role in Development* (pp. 15–28). Hillsdale, NJ: Lawrence Erlbaum.

Mundy, P. (2018). A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *European Journal of Neuroscience*, 47(6), 497–514.

Peacocke, C. (2005). Joint attention: Its nature, reflexivity, and relation to common knowledge. In N. M. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint Attention: Communication and Other Minds. Issues in Philosophy and Psychology* (pp. 298–324). Oxford: Oxford University Press.

Reddy, V. (2010). *How Infants Know Minds*. Cambridge, MA: Harvard University Press.

Roessler, J. (2005). Joint attention and the problem of other minds. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint Attention: Communication and Other Minds. Issues in Philosophy and Psychology* (pp. 230–259). Oxford: Oxford University Press.

Schiffer, S. R. (1972). *Meaning*. Oxford: Clarendon Press.

Schmitz, M. (2015). Joint attention and understanding others. *Synthesis Philosophica*, 29(2), 235–251.

Seemann, A. (2019). *The Shared World: Perceptual Common Knowledge, Demonstrative Communication, and Social Space*. Cambridge, MA: MIT Press.

Siposova, B. & Carpenter, M. (2019). A new look at joint attention and common knowledge. *Cognition*, 189, 260–274.

Sperber, D. & Wilson, D. (1995). *Relevance: Communication and Cognition*. Oxford: Blackwell, 2nd edition.

Thomas, K. A., DeScioli, P., Haque, O. S., & Pinker, S. (2014). The psychology of coordination and common knowledge. *Journal of Personality and Social Psychology*, 107(4), 657–676.

Todd, P. M. & Gigerenzer, G. (2012). What is ecological rationality? In *Ecological Rationality: Intelligence in the World* (pp. 3–30). Oxford: Oxford University Press.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Lawrence Erlbaum.

Tomasello, M. (2003). *Constructing a Language*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.

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

Van Fraassen, B. C. (2002). *The Empirical Stance*. Princeton, NJ: Princeton University Press.

Wilby, M. (2010). The simplicity of mutual knowledge. *Philosophical Explorations*, 13(2), 83–100.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.

Zawidzki, T. W. (2013). *Mindshaping: A New Framework for Understanding Human Social Cognition*. Cambridge, MA: MIT Press.

Chapter 4

Paper III: Coordinating Attention Requires Coordinated Senses

Battich, L., Fairhurst, M., and Deroy, O. (2020). Coordinating attention requires coordinated senses. *Psychonomic Bulletin & Review, 27*(6), 1126–1138. doi: 10.3758/s13423-020-01766-z.

Author contributions:

L.B., M.F. and O.D. conceived of the research idea. L.B. wrote the original draft, and revised the paper with help from all other authors.

Psychonomic Bulletin & Review (2020) 27:1126–1138
https://doi.org/10.3758/s13423-020-01766-z

THEORETICAL REVIEW

Coordinating attention requires coordinated senses

Lucas Battich [1,2] · Merle Fairhurst [1,3,4] · Ophelia Deroy [1,3,5]

Published online: 14 July 2020

Abstract
From playing basketball to ordering at a food counter, we frequently and effortlessly coordinate our attention with others towards a common focus: we look at the ball, or point at a piece of cake. This non-verbal coordination of attention plays a fundamental role in our social lives: it ensures that we refer to the same object, develop a shared language, understand each other's mental states, and coordinate our actions. Models of joint attention generally attribute this accomplishment to gaze coordination. But are visual attentional mechanisms sufficient to achieve joint attention, in all cases? Besides cases where visual information is missing, we show how combining it with other senses can be helpful, and even necessary to certain uses of joint attention. We explain the two ways in which non-visual cues contribute to joint attention: either as enhancers, when they complement gaze and pointing gestures in order to coordinate joint attention on visible objects, or as modality pointers, when joint attention needs to be shifted away from the whole object to one of its properties, say weight or texture. This multisensory approach to joint attention has important implications for social robotics, clinical diagnostics, pedagogy and theoretical debates on the construction of a shared world.

Keywords Joint attention · Social cognition · Cross-modal attention · Multisensory perception

There is more to joint attention than meets the eye

Infant and caregiver coordinate their attention on a toy while learning its name; jazz musicians jointly attend to the music they play together, and hunters can jointly track the smell or sounds of prey in the forest. The ability to coordinate our perception on a shared object of interest comes to most of us between the ages of 9 and 18 months. In our everyday life, we continue to rely on this non-verbal skill, otherwise known as joint attention, to communicate, share experiences, and coordinate with others.

Joint attention has been proposed as one of the essential ingredients of social skills in humans (Adamson, Bakeman, Suma, & Robins, 2019; Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998; Eilan, Hoerl, McCormack, & Roessler, 2005; Moore & Dunham, 1995; Seemann, 2011; Tomasello & Farrar, 1986) and, arguably, across other animal species (Ben Mocha, Mundry, & Pika, 2019; Leavens & Racine, 2009). In most of these accounts, joint attention is measured through the capacity to follow gaze and pointing gestures and coordinate on visible targets (Mundy & Newell, 2007). But does coordinating on visible objects only depend on vision? And what happens when we need to coordinate, not on visible targets, but on auditory, tactile, or multisensory ones?

Uncontroversially, shouting or touching someone's shoulder can be useful to make someone pay attention or orient in the right direction. The role of auditory or tactile alerting signals as accessory cues is well established in primate (Liebal, Waller, Burrows, & Slocombe, 2014) and non-primate (Ben Mocha et al., 2019; Bro-Jørgensen, 2010; Rowe, 1999) animal multimodal communication. It is similarly uncontroversial that non-visual senses often act as a *background* or mere

✉ Lucas Battich
lucas.battich@campus.lmu.de

1 Faculty of Philosophy and Philosophy of Science, Ludwig Maximilian University Munich, Geschwister-Scholl-Platz 1, Munich 80359, Germany

2 Graduate School of Systemic Neurosciences, Ludwig Maximilian University Munich, Munich, Germany

3 Munich Center for Neuroscience, Ludwig Maximilian University Munich, Munich, Germany

4 Institut für Psychologie, Fakultät für Humanwissenschaften, Universität der Bundeswehr München, Munich, Germany

5 Institute of Philosophy, School of Advanced Study, University of London, London, UK

enabling condition for visual attention (for instance, by using vestibular and proprioceptive cues to determine the spatial orientation of one's body in the world, and orient visual attention accordingly). Existing work in the domain of joint attention would certainly accept that other sensory modalities are involved or that joint attention occurs in multisensory settings. Highlighting that joint attention is fundamentally a multisensory phenomenon, however, stresses that non-visual senses are not merely accessories to what could otherwise be defined as a visual phenomenon. Our goal is to provide a more systematic representation of how non-visual sensory resources contribute to joint attention. More specifically, we argue that non-visual senses play two crucial roles. First, they interact closely with gaze and pointing gestures to prime or *enhance* the coordination of visual attention. Non-visual senses can certainly act as distractors, having a negative impact on joint attention. In most cases, however, and with the exception of rare clinical or artificial cases, which we discuss below, other senses are at least minimally involved in the success of joint attention. Second, they play a *necessary* role when it comes to extending social coordination to non-visual and amodal properties of objects and events in the world.

Consider what would happen if gaze and pointing were indeed all there was to the coordination of attention: without computing information from multiple senses, either serially or in conjunction, our referential intentions would run a much higher risk of remaining ambiguous (see *Non-visual senses enhance visual joint attention*). We could not coordinate on non-visible and more abstract aspects of the world (see *Non-visual senses are necessary to extend joint attention*). The current multisensory account is better than a strictly visual one when it comes to explaining how joint attention establishes a socially shared world, where mind-independent objects can be attended in common (see *Theoretical implications: Sharing more than a visual world*). It also has implications for clinical settings and social robotics which are currently focused on gaze-following: with our new account, deficits in gaze coordination could potentially be compensated for by non-visual modalities, and social robots could coordinate attention with humans even without fine-grained gaze-following capacities (see *Applications: Multisensory strategies for the clinic, the school and social robotics*).

Visual joint attention

When Jerome Bruner and colleagues introduced the term *joint attention* to the research on the ontogeny of communication (Bruner, 1974; Scaife & Bruner, 1975), they referred to infants' developing capacity to share their experiences about objects and events with others, and learn word meanings. Now, the construct is used to explain many aspects of our social activities: joint attention in infancy predicts future social competence (Mundy & Sigman, 2015) and emotion regulation, and may reinforce executive functions (Morales, Mundy, Crowson, Neal, & Delgado, 2005; Swingler, Perry, & Calkins, 2015). For adults, engaging in joint attention modulates multiple cognitive abilities (Shteynberg, 2015), including working memory (Gregory & Jackson, 2017; Kim & Mundy, 2012), mental spatial rotation (Böckler, Knoblich, & Sebanz, 2011), and affective appraisals to objects in the environment (Bayliss, Paul, Cannon, & Tipper, 2006).

Bruner's pioneering work centered on joint *visual* attention (Scaife & Bruner, 1975). By and large, subsequent research has remained exclusively focused on the visual domain. Gaze behavior can be easily measured and controlled in laboratory conditions and is therefore a powerful means to study joint attention. In arguing for a multisensory approach, we do not aim to diminish the important role played by gaze cues. Decades of research on gaze following and gaze alternation have firmly established their importance in development and cognition (Flom, Lee, & Muir, 2017; Frischen, Bayliss, & Tipper, 2007; Schilbach, 2015; Shepherd, 2010), and have provided a solid basis for the study of joint attention.

Research into the early development of joint attention distinguishes between *responding* to joint attention by following the direction of others' attention, and *initiating* joint attention by directing or leading the attention of others to a third object or event (Mundy & Newell, 2007). Responding to joint attention, sometimes considered equivalent to following someone's perceptual cues, is the most studied form of joint attention (Fig. 1a) (Mundy, 2018; but see, e.g., Bayliss et al., 2013; Stephenson, Edwards, Howard, & Bayliss 2018). Whether following social cues for attention differs from following non-social cues like arrows remains a topic of debate and investigation, but uncontroversially engages spatial skills and perceptual gaze processing (Gregory, Hermens, Facey, & Hodgson, 2016; Hermens, 2017; Langton, Watt, & Bruce, 2000; Mundy, 2018; Shepherd, 2010). Senses other than vision can play an instrumental role alongside gaze and pointing gestures to guide spatial attention to visible objects.

Attention following, however, is often not sufficient for joint attention. For example, I can follow your attention without you noticing in any way that I did so, which would not count as joint attention. In addition to gaze following, joint attention requires the ability to engage in a *reciprocal coordination* that guarantees we are looking at the same object together (Mundy, 2018; Siposova & Carpenter, 2019) (Fig. 1b). This triadic coordination exhibits the understanding, even minimally, that both agents are mutually aiming at or aware of the object (Bakeman & Adamson, 1984; Mundy, 2016; Tomasello, 1995). Non-visual senses here may do more than facilitate attention following: they help to strategically select the appropriate target of joint attention between two individuals.

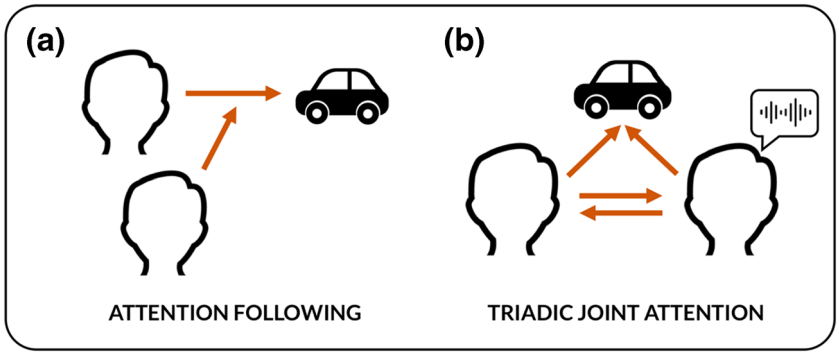

Fig. 1 Following attention is different from coordinating attention. (**a**) Attention following is characterized by the unilateral response of one individual. It can consist of behaviors such as gaze following, or the monitoring of others' bodily posture and gestures, and responding to vocal and haptic cues. Attention following is a pre-condition for full joint attention, and occurs earlier in development. (**b**) Coordination of attention is characterized by the reciprocal interaction between individuals toward a third object. In addition to gaze following, joint attention includes gaze-alternation and directing other's gaze through pointing — but also other senses

Engaging in joint attention requires one to know what one is attending to, as well as what the other is attending to. This in turn requires the combined processing of three types of information: (1) information about one's own attentional state, including interoceptive and proprioceptive information (Mundy & Jarrold, 2010); (2) information about the other's attentional state; (3) information about the target of joint attention (Mundy, 2018; Siposova & Carpenter, 2019). All three types of information and their processing can engage multiple senses, besides vision. Information about my own attention to the object of common reference may include whether I am actively handling the object, or merely looking at it. Information about the other's attentional state will vary depending on whether they have access to the same sensory information I have. The strategies used to establish joint attention will vary when we coordinate on a smell, a sound, the color of an object, or a whole, complex multisensory event.

Non-visual senses enhance visual joint attention

Visual cues provide multisensory expectations

When processing information about the other's attentional state, we can further distinguish between the sense I rely on to monitor the other's attention (e.g., I *gaze* at your hand grasping), and the sense they use, which I monitor to gather information about their attention (e.g., I gaze at your *hand grasping*). This distinction already pleads for the incorporation of richer sensory measures in models of joint attention than mutual eye contact, gaze following or gaze alternation. Observing someone's touching actions, as well as someone being touched, activates similar neural circuits normally involved in the execution of those actions, and the processing of actual touch (Buccino et al., 2001; Keysers et al., 2004), suggesting that tactile expectations regarding the jointly attended object can be gathered vicariously even by sight alone. Studies have here looked at the use of coupled information from eye and hand gestures. When reaching and manipulating objects, gaze and hand movements are systematically coordinated with respect to the target object, with gaze fixation leading the subsequent hand movement (Horstmann & Hoffmann, 2005; Pelz, Hayhoe, & Loeber, 2001). This eye-hand coupling can provide a path for well-coordinated rapid and successful joint attentional interaction: although gaze provides a faster cue to the spatial area where the target is located, the hand trajectory while reaching and grasping provides a slower but more spatially precise and stable cue to the target's location (Yu & Smith, 2013). Additionally, in following a grasping gesture, observers are sensitive to both the direction and the grip aperture size of the reaching hand to facilitate target detection (Tschentscher & Fischer, 2008). Reliance on multiple senses and their interaction may here help provide richer spatial and temporal representations of our environment (Keetels & Vroomen, 2012; Stoep, Postma, & Nijboer, 2017). These multisensory strategies are present during infant–caregiver joint attentional engagement, which reflects the multisensory nature of parent–infant dyadic communication (Gogate, Bahrick, & Watson, 2000; Gogate, Bolzani, & Betancourt, 2006; Hyde, Flom, & Porter, 2016). Multimodal behaviors help sustain joint attention between parents and infants from 12 to 16 months old, in particular when parents express some interest in an object looking at, talking about, and touching the jointly attended object (Suarez-Rivera, Smith, & Yu, 2019). One-year-old infants do not tend to follow the partner's gaze to monitor their attention while playing together with a toy. Instead, they follow their hands (Yu & Smith, 2013). Taken together, this evidence suggests that non-visual senses and multisensory expectations are exploited in joint attention, especially to narrow down the spatial location of the target of joint attention through spatial redundancy.

Recent research on the emergence of pointing gestures reinforces this suggestion. Children interpret pointing gestures

as if they were attempts to touch things (O'Madagain, Kachel, & Strickland, 2019), indicating that understanding visual cues about someone's touch toward a third object are ontogenetically prior to the understanding referential pointing gestures. This recent work suggests new methods to explore whether a similar relation is present in the phylogeny of grasping and pointing cues.

Non-visual cues enhance visual target detection

Joint attention can be established through gaze alone (Flom et al., 2017). In many social contexts, the use of visual cues can be sufficient to coordinate attention, but may not always be the most *efficient*. In information theory, adding redundancy to the initial message so that several portions of the message carry the same information increases the chance that the message is accurately received at the end of a noisy channel (Shannon, 1948). This is also true in perception. For an everyday illustration, consider trying to hit a nail with a hammer. It is possible to push the pointy part of the nail in the wall and then hammer it while relying only on vision, but by holding the nail with one hand, you can gather information about the nail's spatial position both through vision and through your hand position. Studies in multisensory perception demonstrate that redundant information delivered across several sensory modalities increases the reliability of a sensory estimate: it enhances a perceiver's accuracy and response time to detect the presence of a stimulus and to discriminate and identify a sensory feature (e.g., an object's shape or its spatial location), a so-called *redundant-signals effect* (Ernst & Banks, 2002; Miller, 1982). It is safe to assume that redundancy of information across modalities is also usefully exploited when establishing and sustaining joint attention. For example, the caregiver will point to a toy car that the infant can see, and tap on the toy to make a noise. Here, the combination of the visual and auditory information enhances the infant's accuracy and speed in shifting spatial attention (cf. Partan & Marler, 1999) (see Fig. 2a). In this section we review how multisensory information facilitates visual coordination and target detection, focusing on three mechanisms: spatial congruency, temporal synchrony, and cross-modal correspondences.

Redundancy of spatial information is shown to help with the orienting of visual attention in experiments where individual perceivers are presented with a task-irrelevant cue on the same or opposite side of the subsequent visual target. Participants tend to respond more rapidly, and more correctly, to visual targets appearing at the same location as the former task-irrelevant cue, rather than on the opposite side. This works for visual irrelevant cues (Posner, 1980; see Carrasco, 2011; Wright & Ward, 2008, for overviews) and also occurs across modalities: participants are faster and more accurate at detecting target stimuli in one modality when a task-irrelevant cue is presented in the same or similar location (McDonald, Teder-Sälejärvi, & Hillyard, 2000; Spence, McDonald, & Driver 2004a; see Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010, for a review). This evidence suggests that when participants direct their spatial attention to a certain location driven by one modality, their sensitivity to stimuli in that location is also enhanced for other modalities. While these traditional cross-modal attention studies use nonsocial stimuli, there is growing evidence of similar effects with social ones. Gaze-cueing experiments using covert orienting paradigms have shown that cues from another's gaze behavior facilitate the processing of tactile stimuli at the body location corresponding to the other's gaze direction (Soto-Faraco, Sinnett, Alsius, & Kingstone, 2005). Recent work shows that gaze-based cues enhance the processing of tactile (De Jong & Dijkerman, 2019) and auditory (Nuku & Bekkering, 2010) stimuli at what is meant to be the jointly attended location. The current evidence of cross-modal effects in spatial attention gives us reason to think that a wide array of sensory cues, besides someone's gaze or gesture direction, can be exploited to assist spatial coordination between joint attenders.

Temporal synchrony between cross-modal cues, in the absence of spatial congruence, also directs someone's spatial attention. Van der Burg et al. (2008, 2009, 2010) have shown that the presentation of a spatially irrelevant cue in the auditory or tactile modality can facilitate a participant's visual search performance in an environment with color-changing elements, when the non-visual cue is presented at the same time as a color change in the target element. Known as the "pip-and-pop effect," these studies show that even when one sensory cue does not carry relevant spatial information, it can enhance the salience of a spatially relevant cue in a different modality (Ngo & Spence 2010). These cross-modal effects could be exploited in trying to establish joint attention to a target in a changing, dynamic environment. Touching someone's shoulder or vocalizing in synchrony with a certain movement or event (e.g., every time a particular bird jumps from a branch or flutters its wings) may be a better strategy to coordinate attention to it than pointing alone (Fig. 2B).

Finally, *the properties of the non-visual social cues* can also shape congruency effects, besides providing spatial or temporal congruence with visual cues. We are not talking here of semantic congruence (saying "dog" or "woof" while pointing at the visible dog) but of sensory congruence between properties such as pitch or loudness, and visual properties, such as brightness, shape, etc. Humans, like some other animals (Bee, Perrill, & Owen, 2000), exploit the environmental regularities that exist between sensory cues across modalities for communicative purposes. Such regularities show up in cross-modal correspondences, i.e. robust associations between independent features or dimensions across modalities (Spence, 2011; Spence & Deroy, 2013). For example, high-pitched sounds correspond to high spatial positions of a visual stimulus, so that when both features are congruently matched,

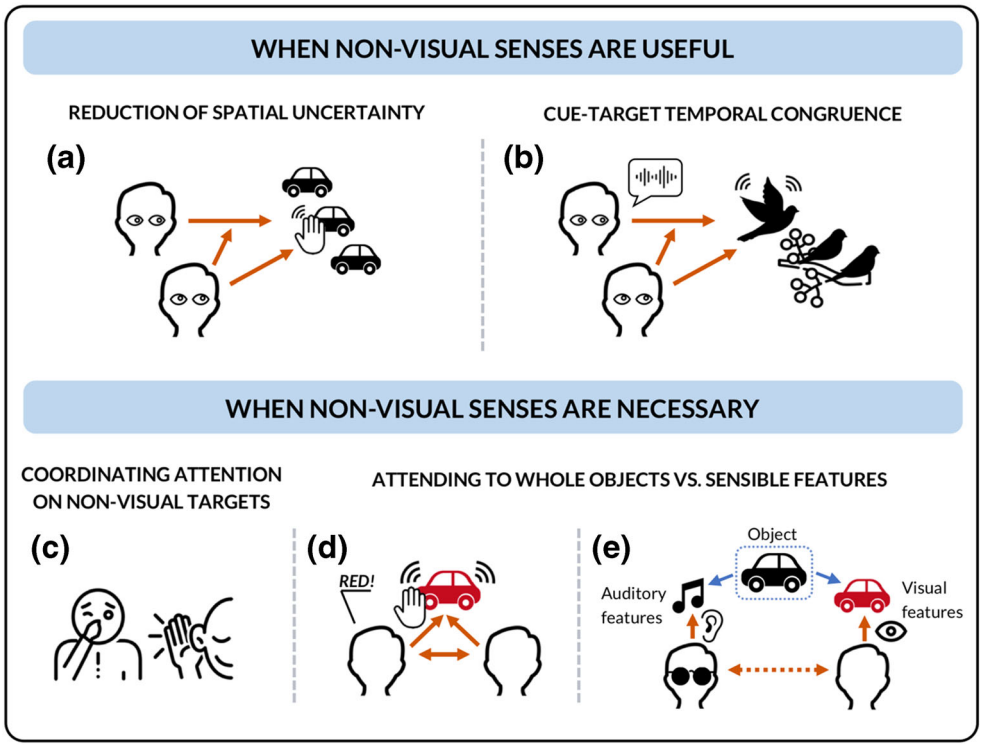

Fig. 2 (**Upper panel**) Non-visual cues can complement visual cues in joint attention. (**a**) Redundant information delivered across modalities can increase accuracy and speed in following spatial cues: by monitoring someone's eye-gaze cues in combination to their hand-grasping actions, the follower's response in localizing the object of joint attention is enhanced. (**b**) Using temporal congruence between a cue and a target in different modalities to facilitate someone's orienting to the correct visual target. (**Lower panel**) Non-visual cues are often necessary for joint attention. (**c**) Establishing joint attention toward a non-visual target by using ostensive visual cues: ostensive pointing at the relevant sensory organ (touching one's ear or one's nose) can provide evidence to another agent of the intention of attending to a non-visual stimulus (a sound, a smell). Such strategies rely on cognitive abilities to infer that the target is non-visual. (**d**) Exploiting temporal synchrony: a parent shakes an object in a temporally synchronous manner congruent with their uttering the word "red." While the visual stimulus and the auditory stimulus have different causal sources (the toy and the parent), the information is conveyed that the word "red" is associated with a visual property of the toy. (**e**) Coordinating on objects we each experience through different modalities: each subject must process information about each other's modal access relative to the target to successfully achieve coordination

attentional orienting to a target visual cue is facilitated (Bernstein & Edelstein, 1971). Other cross-modal correspondences, such as the one that exists between pitch and brightness, work together with temporal synchrony to elicit a "pip-and-pop effect" during visual search: when a visual target changes brightness, a congruent change in pitch of a task-irrelevant auditory cue enhances correct target detection (Klapetek, Ngo, & Spence 2012). The effects of cross-modal correspondence have so far been mostly studied in nonsocial domains. We suggest that they are also relevant in social domains. For example, when trying to direct your attention to an animal hiding in the trees, emitting a high-pitched rather than a low-pitched interjection might help direct attention to the higher part of the scene. To test this suggestion, future work on multisensory joint attention will have to address the role of cross-modal alerting signals, and how the processing of cross-modal social signals compares to nonsocial situations.

Importantly, how much spatial, temporal, and cross-modal congruence facilitate the processing of visual gaze or pointing gestures is ripe for more precise measurements, notably by artificially manipulating the discrepancy between the cues, and measuring the subsequent effects on joint attention.

The interplay between coordinated attention and multisensory processing

Multisensory cues can help the social coordination of attention. Surprisingly, the reverse can also be true. A few innovative studies give evidence that coordinating attention with a partner modulates a participant's multisensory processing. People are better able to ignore task-irrelevant stimuli in a distracting modality when they know that someone else is attending to these distractors (Heed, Habets, Sebanz, & Knoblich, 2010; Wahn, Keshava, Sinnett, Kingstone, & König, 2017).

In the first study (Heed et al., 2010), participants had to judge whether a tactile stimulus was presented on the upper or lower part of a cube, while a distractor visual stimulus was presented synchronously at the same or opposite elevation. In the individual task, participants responded faster and more

accurately when the distractor stimulus was presented at the same elevation as the tactile target, showing a performance difference known as the cross-modal congruency effect (CCE; see Spence, Pavani, Maravita, & Holmes, 2004b, for a review). Interestingly, the CCE was significantly reduced when a partner was instructed to attend to the visual stimuli, indicating that participants could better ignore incongruent distractors when their partner responded on them. This effect was recently replicated in an audiovisual congruency task (Wahn et al., 2017) involving visual flashes and auditory tones originating from the same or opposite spatial vertical location. Knowing that someone else was attending to the incongruent flashes allowed participants to respond faster to the tones, resulting in a reduced CCE.

These studies show that responding jointly reduces the interference of competing stimuli in a multisensory setting (Wahn & König, 2017). The results seem at odds with a recent tradition of research showing that acting jointly increases the interference of irrelevant stimuli, presumably due participants co-representing each other's tasks besides their own (Sebanz, Knoblich, & Prinz, 2003, 2005). For example, performing an object-based visual attention task jointly impairs performance (Böckler, Knoblich, & Sebanz, 2012), and the increase in interference of irrelevant information is well documented in Go/No Go joint Simon tasks (Dolk et al., 2014). The difference between the reduction and the increase of irrelevant interference in different joint attentional tasks may be due to the nature of the tasks studied. An efficient division of labor can be allowed when the different target stimuli of each co-actor's task are presented concurrently, whereas the beneficial effect of filtering irrelevant information disappears when the task involves two competing Go/No Go actions (Dolk & Liepelt, 2018; Sellaro, Treccani, & Cubelli, 2018).

So far, studies have focused on coordinated social attention to separate cross-modal targets. Each participant attends and responds to a different modal stimulus, which facilitates a perceptual division of labor. A multisensory approach to joint attention should encourage us to extend this work to situations where partners attend and respond to the same multisensory stimuli, or try and ignore distractors in the same modality while focusing on another one. For example, when two subjects jointly coordinate their attention toward sounds and flashes presented closely in space and time, the binding of two or more modal features may be further enhanced, compared to conditions where subjects attend to the same sounds and flashes alone. If both are asked to attend jointly to the sounds, and jointly ignore the flashes, they may also be less prone to a ventriloquist effect, where the location of the sounds is displaced toward the location of the flashes (Vroomen & De Gelder, 2004).

Non-visual senses are necessary to extend joint attention

Jointly attending to invisible sounds or smells

The dominance of vision in the study of, and theorizing about, perception and joint attention may reflect the importance of this modality in humans (Colavita, 1974; Emery, 2000; Itier & Batty, 2009; Sinnett, Spence, & Soto-Faraco, 2007), but should not occult the fact that humans also jointly attend and teach words referring to sounds and smells, not to mention musical features.

Establishing joint attention toward a non-visual target requires access to information about both the other's attentional focus and, crucially, the target where the other's attention should be directed. Relative to gaze, a clear limitation of audition and olfaction is that their target of attention is not publicly disclosed to an observer. To establish joint attention coordination on strictly non-visual targets, subjects may be obliged to *indirectly* coordinate on the visual location of these non-visual events and use cognitive strategies to signal and to infer that the target is non-visual. For example, ostensive pointing at the relevant sensory organ (touching one's ear, or one's nose) can provide evidence to another agent of the intention of attending to a non-visual stimulus (Baker & Hacker, 2005) (Fig. 2c).

In addition, ostensive strategies could involve *negative* cues such as standing still, and keeping one's head and eyes motionless to signal that attention should be directed to a non-visual target of joint attention. Here, one prediction would be that such cases would occur only *after* expectations about pointing and gaze have been fully formed – as the strategy rests on using a mismatch between the expectation (that eyes and heads move) and the results (eyes and heads do not move, meaning that the target is non-visible).

Although visual and gestural ostensive cues may be used on some occasions to direct attention to a non-visible target, such behaviors already presuppose that the other agent is capable of understanding that sounds and smells are objects in the world that can be perceived together with others. The developmental onset of the ability to gaze at objects jointly with others is well researched. One outstanding question is when infants start to display an equivalent understanding that others can share with them attention to smells and sounds, and how this understanding is coupled with processing the visual attention of others.

Jointly attending to amodal features

Gaze-based joint attention enhances basic object recognition, even in very young infants (Cleveland & Striano, 2007; Hoehl, Wahl, Michel, & Striano, 2012; Wahl, Marinović, & Träuble, 2019). However, object-recognition development

relies on the ability to perceive global, invariant, and amodal properties like spatial location, tempo, rhythm, and intensity, which can only be conveyed through the combination of different sense modalities (Bahrick & Lickliter, 2014; Hyde et al., 2016). The redundancy introduced by multisensory events can thus be strategically used to establish joint attention on amodal features of objects and events. Bahrick and colleagues suggest that perception of this amodal information is critically important for the development and performance of perceptual object and event recognition (Bahrick, 2010).

One key example of such strategic use is the manner in which the temporal synchrony (when onset, offset, and/or duration of sensory stimuli are the same) between vision and audition can be exploited. For instance, a parent will shake an object in a temporally congruent way with the word they utter, thus enhancing the associating between object and word (Fig. 2d) (Gogate et al., 2000; Gogate & Hollich, 2016; Jesse & Johnson, 2016). The significance of temporal and spatial synchrony across different sensory cues is not only restricted to language learning. Running a toy car over the table or over the infant's arm while saying "vroom" may not directly lead towards word acquisition, as there is no linguistic element to be acquired. But it may help to bind both visual (e.g., shape) and auditory (e.g., vehicle noises) properties to the same object, the toy car.

The use of two cues highlights an important point. Here the target of joint attention is broader than the cues used to attract and coordinate attention: making a sound while moving a toy-car and looking at it ostensibly can be used to draw attention to the whole multisensory object, including its amodal extension, its weight, texture, etc., and not just its auditory or visual properties.

Conversely, the target of joint attention can be narrower than the object of individual attention and even of mutually shared experiences. For example, while musicians may attend to how others move their bow, hands, and heads, their joint attention is focused on the music they produce or, indeed, an element of the music (a particular voice or a particular theme). Moreover, their auditory joint attention will be coordinated through the gestures of a musical conductor, which provide visual cues about particular aspects in the sounds that musicians must follow – the music's tempo, for example. In this sense, the target of coordinated attention is narrower than the visual and auditory cues they use to attract and maintain their attention and narrower than the multisensory production that they know they are mutually experiencing.

Taking into account the role of non-visual senses in coordinating attention highlights that the *target of joint attention* can often be different than *the target of each individual's attention*. Joint attention involves more than merely orienting toward the same target. Perceptual attention can be characterized as the selective information processing of a specific area or features of the sensory world, while ignoring or decreasing processing of other areas and features (Eriksen & James, 1986; Klein & Lawrence, 2012). Joint attention results in a socially mediated enhancement in the processing of sensory information (Mundy, 2018). In other words, joint attention brings about another level of selectivity over an individual's own perceptual attention. Engaging in joint attention allows us to extract from a fundamentally multisensory experience the relevant integrated targets or specific features (visual, auditory, etc.) for further information processing and social coordination.

Sensory deficits: Jointly attending to a multisensory object through different senses

What happens when coordination occurs on objects that the two agents experience through different modalities? This is the case when coordinating attention with blind individuals, or individuals whose vision is temporarily blocked (say, they wear opaque glasses). Here, both or at least one agent knows that the other cannot access the object on which attention needs to be coordinated via the visual modality that they themselves use to access the object.

Cases of sensory deprivation (e.g., deafness, blindness, anosmia, hyposmia) provide methodological tools to study the roles of different senses during joint attention, and how individuals with limited sensory access negotiate coordination. Atypical development highlights the manner in which we share attention with others as a function of information access. In a case study of two congenital blind infants, coordinating attention with their caregivers involved auditory information as well as tactile and kinesthetic information, memory, sound changes, air currents, and echolocation (Bigelow, 2003). Deaf-blind children tend to combine two or more sensory sources for coordinating attention toward an object with their non-deaf-blind parents (Núñez, 2014). A 3-year-old child with profound visual and hearing impairment would first draw on touch to check that she has her caregiver's attention. She would then hold the object of interest towards the caregiver's face with one hand while continuing to monitor their attention with the other hand, vocalizing excitedly and smiling throughout (Núñez, 2014). Social gaze behavior and joint attention through vision alone can also be impacted by auditory deficits (e.g., Corina & Singleton, 2009; Lieberman, Hatrak, & Mayberry, 2014). There is evidence, for example, that auditory deprivation affects the effect of gaze cues and gaze following. Deaf children (aged between 7 and 14 years old) are more susceptible to the influence of task-irrelevant gaze cues than hearing children (Pavani, Venturini, Baruffaldi, Caselli, & van Zoest, 2019). This effect appears to dissipate in deaf adults, suggesting that the salience of social gaze cues changes during development (Heimler et al., 2015)

These studies reinforce the view that our ability to establish the triadic relation characteristic of joint attention can vary

according to the modal pathways used for directing and following the other's attention (Fig. 2e). In multisensory contexts, agents can share across information to which the other person has no access, or is not actively accessing. To illustrate, suppose we are jointly attending to a coffee cup by vision. In addition, I am also touching the object to judge its temperature. Through our coordinated attention to the cup and by monitoring my responses, you can vicariously gather information on my haptic experience and whether the cup is warm.

Theoretical implications: Sharing more than a visual world

Philosophers and psychologists have taken the role of joint attention in our understanding of other minds to argue that joint attention is, in fact, essential to understand the concept of a shared objective world, where mind-independent objects are attended in common (Davidson, 1999; Eilan, 2005; Engelland, 2014; Seemann, 2019; Tomasello, 2014). The ability to coordinate attention to an object together with another individual goes hand in hand with the ability to experience the object as a mind-independent entity separate from oneself (Campbell, 2011). This view has pre-eminent precursors in psychology. Lev Vygotsky (2012), in particular, held the doctrine that all higher cognition in an individual arises from an internationalization process of prior social interactions. Vygotsky's original formulation may seem overly strong, but a Vygotskyan approach has become increasingly influential to account for the social influences observed in the development of cognition and psychiatric disorders (Bolis & Schilbach, 2018; Fernyhough, 2008; Hobson & Hobson, 2011; Tomasello, 2019). Granting that joint attention helps us build a shared objective world, restricting ourselves to gaze and vision alone would make this world incredibly impoverished.

To stress this point, imagine a case where joint attention would *only* occur through gaze-following and looking at pointing gestures: we would only be able to coordinate attention on the visual properties of objects and events. We would certainly be able to learn that most bananas are yellow; we would learn that using color-tinged glasses changes how these properties look; and we would learn that other people may be seeing a drawing upside down when we see it right side up. But how would two people jointly attend to the sound of thunder, or the smell of natural gas? Would they quickly make the difference between pointing at the color of the car, or the car as a whole?

Realizing that we attend to a unitary object or to specific properties cannot occur in a visual-only scenario, or certainly without resorting to more conventional or linguistic means. Using a multisensory combination of cues is necessary to explain that we share an objective world of multisensory objects, sounds, smells, and textures.

Applications: Multisensory strategies for the clinic, the school, and social robotics

A better understanding of the mechanisms through which multisensory and cross-modal processes help and shape the successful coordination of attention on the same object, or on a given aspect of an object, can have direct implications for several sectors and fields.

When gaze coordination is limited

In a caregiver-child pair in which one person has a sensory deficit (deaf-blind, deaf, blind), the information that can be shared will be limited in some way, and compensated for in others. Tactile joint attention is crucial for children with visual impairments and multiple sensory disabilities (Chen & Downing, 2006). A child rolling Play-Doh will lead the adult's hand to share attention to her activity. The adult can follow the child's lead and focus on what the child is doing by keeping non-controlling tactile contact both with the child's hands and with the Play-Doh, establishing a reciprocal relation.

An emphasis on gaze interaction, however, can lead to biased assessments of an individual's ability to coordinate and interact with others. When measured according to vision-based operationalizations, deaf children of hearing parents show a delay in the onset of *visual* joint attentional skills, and symbol-infused joint attention (involving words or symbolic gestures) tends to be less frequent than in typically developing infants (Prezbindowski, Adamson, & Lederberg, 1998). These results have been challenged when factoring the role of other senses: hearing parents do accommodate their deaf children's hearing status by engaging them via multiple modalities, while parents of typically developing children tend to use alternating unimodal (either visual or auditory) cues during a joint attention episode (Depowski, Abaya, Oghalai, & Bortfeld, 2015). Developmental differences are not pronounced in deaf children of deaf parents, who tend to coordinate attention using both visual and tactile signals (Spencer, 2000).

Taken together, these findings suggest that operationalizations of joint attention based on gaze alone may produce unreliable measures of the real ability of infants to coordinate attention with others. They also show that non-visual senses impinge on the development of joint attention, even for non-visually impaired deaf individuals. Finally, the ability to engage in joint attention depends not just on the atypical infant's behavior, but, importantly, on that of their caregivers. Adopting a multisensory perspective on joint attention can provide better measures of the development of atypical children and inspire new complementary strategies to foster the development of joint attention skills.

Multisensory joint attention during learning

The ostensive character of joint attention is central to the acquisition of language (Adamson et al., 2019; Carpenter et al., 1998; Tomasello & Farrar, 1986) and, more generally to the transmission of knowledge and learning (Csibra & Gergely, 2009). In traditional paradigms on the role of joint attention in language development, triadic coordination to a target object is visually established through gaze alternation or pointing, accompanied by the utterance of the linguistic label to be associated with the object (see Akhtar & Gernsbacher, 2007, for a critical overview). As noted above, however, early linguistic development is increasingly recognized as a multisensory process (Gogate & Hollich, 2016; Jesse & Johnson, 2016). Similarly, the importance of multisensory teaching methods is increasingly recognized within pedagogy, both for typically developing children (e.g., Kirkham, Rea, Osborne, White, & Mareschal, 2019; Shams & Seitz, 2008; Volpe & Gori 2019) and for children with learning differences, including dyslexia (e.g., Birsh, 2005) and autistic spectrum disorder (e.g., Mason, Goldstein, & Schwade, 2019).

A better understanding of the interplay of different sense modalities during joint attention, across different ages and neurological conditions, can support the development of multisensory protocols in pedagogical situations. It should also be a reminder of cross-cultural differences when generalizing about teaching: in some cultures, touch, sounds, or smells are more central to social engagement, learning, or communication (Akhtar & Gernsbacher, 2008; Kinard & Watson, 2015). Akhtar and Gernsbacher (2008) review evidence suggesting that in cultures where infants experience continuous physical or vocal contact with their caregivers, and spend less time in face-to-face eye contact, evidence of social engagement will rely on tactile, auditory, and olfactory cues more than mutual gaze cues. Mothers in Kenya, for example, engage in more touching and holding with their infants, and less in eye contact, than mothers in the USA (Richman, Miller, & LeVine, 1992).

Multisensory joint attention with artificial social agents

The field of social robotics strives to bring artificial agents into hospitals, schools, businesses, and homes – complex social environments that require the enactment of naturalistic nonverbal interactions, including joint attention coordination (Clabaugh & Matarić, 2018; Kaplan & Hafner, 2006; Yang et al., 2018). For a robot to help a human partner assemble a piece of furniture, stack blocks with children in the playground, and assist people with disabilities in their daily lives, they need to be sensitive to what the human is attending to, and asking them to attend to.

Whether an artificial agent can successfully engage in joint attention with humans will depend on how well they can meet the behavioral expectations of their human interaction partner. Will they be able to both initiate and follow attentional cues in a naturalistic manner (Pfeiffer-Leßmann, Pfeiffer, & Wachsmuth, 2012)? One current approach is to enable social robots to mimic human gaze behaviors (Admoni & Scassellati, 2017; Kompatsiari, Ciardo, Tikhanoff, Metta, & Wykowska, 2019). However, while human participants do respond to the gaze of artificial agents (Willemse, Marchesi, & Wykowska, 2018), they are also highly sensitive to momentary multimodal behaviors produced by their artificial partner (Yu, Schermerhorn, & Scheutz, 2012). By adopting a multisensory perspective on human-robot joint attention, it is possible to examine non-visual cues emitted by the artificial agent, so that they accord with the expectations of human interaction partners. Being sensitive to the non-visual cues emitted by humans could also improve the spatial and temporal resolution of attention-orienting in robots.

Conclusion

Any episode of visual attention will, de facto, rely on background multisensory processing: we rely on proprioceptive and vestibular cues to visually orient our attention and ourselves in the world. Multisensory interactions, however, play a more substantial role in the coordination of attention across social agents: infants and adults recruit multiple sense modalities to initiate and follow someone's attention to a specific object or location in space. These interactions can be distinguished depending on whether they facilitate the coordination of visual attention, or whether they extend the coordination to non-visual and amodal properties. While non-visual modalities are useful complements for vision in the former case, they are essential in the latter case: some kinds of joint attention are necessarily multisensory, and could not be carried by vision alone.

This multisensory approach has implications for behavioral and developmental models of joint attention. Just as selective attention can be described as a cognitive capacity that both influences and is influenced by perceptual processes across different modalities, models of joint attention must be flexible enough to incorporate how it relies on dynamic information from multiple senses. It also has practical implications to overcome clinical deficits in joint attention, augment its pedagogical role, and address the challenge of coordinating attention between humans and social robots.

Author Note MF was supported in part by funds from LMU Munich's Institutional Strategy LMUexcellent within the framework of the German Excellence Initiative. OD is supported by the NOMIS foundation "Dise" grant.

Funding Information Open Access funding provided by Projekt DEAL.

Compliance with ethical standards

Conflicts of interest The authors have no conflicts of interest to disclose.

References

Adamson, L. B., Bakeman, R., Suma, K., & Robins, D. L. (2019). An expanded view of joint attention: Skill, engagement, and language in typical development and autism. *Child Development*, *90*(1), e1–e18. https://doi.org/10.1111/cdev.12973

Admoni, H., & Scassellati, B. (2017). Social eye gaze in human-robot interaction: A review. *Journal of Human-Robot Interaction*, *6*(1), 25–63. https://doi.org/10.5898/JHRI.6.1.Admoni

Akhtar, N., & Gernsbacher, M. A. (2007). Joint attention and vocabulary development: A critical look. *Language and Linguistics Compass, 1*(3), 195–207. https://doi.org/10.1111/j.1749-818X.2007.00014.x

Akhtar, N., & Gernsbacher, M. A. (2008). On privileging the role of gaze in infant social cognition. *Child Development Perspectives*, *2*(2), 59–65. https://doi.org/10.1111/j.1750-8606.2008.00044.x

Bahrick, L. E. (2010). Intermodal perception and selective attention to intersensory redundancy: Implications for typical social development and autism. In *The Wiley-Blackwell handbook of infant development* (pp. 120–166). https://doi.org/10.1002/9781444327564.ch4

Bahrick, L. E., & Lickliter, R. (2014). Learning to attend selectively: The dual role of intersensory redundancy. *Current Directions in Psychological Science, 23*(6), 414–420. https://doi.org/10.1177/0963721414549187

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, *55*(4), 1278–1289.

Baker, G. P., & Hacker, P. M. S. (2005). Ostensive definition and its ramifications. In *Wittgenstein: Understanding and meaning. Part i: Essays* (pp. 81–106). https://doi.org/10.1002/9780470752807.ch5

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. https://doi.org/10.3758/BF03213926

Bayliss, A. P., Murphy, E., Naughtin, C. K., Kritikos, A., Schilbach, L., & Becker, S. I. (2013). "Gaze leading": Initiating simulated joint attention influences eye movements and choice behavior. *Journal of Experimental Psychology: General, 142*(1), 76–92. https://doi.org/10.1037/a0029286

Bee, M. A., Perrill, S. A., & Owen, P. C. (2000). Male green frogs lower the pitch of acoustic signals in defense of territories: A possible dishonest signal of size? *Behavioral Ecology*, *11*(2), 169–177. https://doi.org/10.1093/beheco/11.2.169

Ben Mocha, Y., Mundry, R., & Pika, S. (2019). Joint attention skills in wild Arabian babblers (Turdoides squamiceps): A consequence of cooperative breeding? *Proceedings of the Royal Society B: Biological Sciences*, *286*(1900), 20190147. https://doi.org/10.1098/rspb.2019.0147

Bernstein, I. H., & Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology*, *87*(2), 241–247. https://doi.org/10.1037/h0030524

Bigelow, A. E. (2003). The development of joint attention in blind infants. *Development and Psychopathology*, *15*(2), 259–275.

Birsh, J. R. (2005). *Multisensory teaching of basic language skills*. Baltimore: Paul Brookes Publishing Co.

Böckler, A., Knoblich, G., & Sebanz, N. (2011). Giving a helping hand: effects of joint attention on mental rotation of body parts. *Experimental Brain Research*, *211*(3-4), 531–545. https://doi.org/10.1007/s00221-011-2625-z

Böckler, A., Knoblich, G., & Sebanz, N. (2012). Effects of a coactor's focus of attention on task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1404–1415. https://doi.org/10.1037/a0027523

Bolis, D., & Schilbach, L. (2018). 'I interact therefore I am': The self as a historical product of dialectical attunement. *Topoi*, 1–14. https://doi.org/10.1007/s11245-018-9574-0

Bro-Jørgensen, J. (2010). Dynamics of multiple signalling systems: animal communication in a world in flux. *Trends in Ecology & Evolution*, *25*(5), 292–300. https://doi.org/10.1016/J.TREE.2009.11.003

Bruner, J. S. (1974). From communication to language: A psychological perspective. *Cognition*, *3*(3), 255–287. https://doi.org/10.1016/0010-0277(74)90012-2

Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G. & Freund, H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience, 13*(2), 400–404. https://doi.org/10.1111/j.1460-9568.2001.01385.x

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. 415–430). Cambridge, MA: MIT Press.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63*(4), 1–174. https://doi.org/10.2307/1166214

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Chen, D., & Downing, J. E. (2006). *Tactile strategies for children who have visual impairments and multiple disabilities: Promoting communication and learning skills*. New York, NY: AFB Press.

Clabaugh, C., & Matarić, M. (2018). Robots for the people, by the people: Personalizing human-machine interaction. *Science Robotics*, *3*(21), eaat7451. https://doi.org/10.1126/scirobotics.aat7451

Cleveland, A., & Striano, T. (2007). The effects of joint attention on object processing in 4- and 9-month-old infants. *Infant Behavior and Development*, *30*(3), 499–504. https://doi.org/10.1016/J.INFBEH.2006.10.009

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, *16*(2), 409–412. https://doi.org/10.3758/BF03203962

Corina, D., & Singleton, J. (2009). Developmental social cognitive neuroscience: Insights from deafness. *Child Development, 80*(4), 952–967. https://doi.org/10.1111/j.1467-8624.2009.01310.x

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Davidson, D. (1999). The emergence of thought. *Erkenntnis*, *51*(1), 511–521. https://doi.org/10.1023/A:1005564223855

De Jong, M. C., & Dijkerman, H. C. (2019). The influence of joint attention and partner trustworthiness on cross-modal sensory cueing. *Cortex*, *119*, 1–11. https://doi.org/10.1016/j.cortex.2019.04.005

Depowski, N., Abaya, H., Oghalai, J., & Bortfeld, H. (2015). Modality use in joint attention between hearing parents and deaf children. *Frontiers in Psychology*, *6*, 1556. https://doi.org/10.3389/fpsyg.2015.01556

Dolk, T., & Liepelt, R. (2018). The multimodal go-nogo Simon effect: Signifying the relevance of stimulus features in the go-nogo Simon paradigm impacts event representations and task performance. *Frontiers in Psychology*, *9*, 2011. https://doi.org/10.3389/fpsyg.2018.02011

Dolk, T., Hommel, B., Colzato, L. S., Schütz-Bosbach, S., Prinz, W., & Liepelt, R. (2014). The joint Simon effect: A review and theoretical integration. *Frontiers in Psychology*, *5*, 974. https://doi.org/10.3389/fpsyg.2014.00974

Eilan, N. (2005). Joint attention, communication, and mind. In N. Elian, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds. Issues in philosophy and psychology* (pp. 1–33). Oxford: Oxford University Press.

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

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, *24*(6), 581–604.

Engelland, C. (2014). *Ostension*. https://doi.org/10.7551/mitpress/9780262028097.001.0001

Eriksen, C. W., & James, J. D. S. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*(4), 225–240. https://doi.org/10.3758/BF03211502

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Fernyhough, C. (2008). Getting Vygotskian about theory of mind: Mediation, dialogue, and the development of social understanding. *Developmental Review*, *28*(2), 225–262. https://doi.org/10.1016/j.dr.2007.03.001

Flom, R., Lee, K., & Muir, D. (Eds.). (2017). *Gaze-following: Its development and significance*. New York: Psychology Press.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*(4), 694–724. https://doi.org/10.1037/0033-2909.133.4.694

Gogate, L. J., & Hollich, G. (2016). Early verb-action and noun-object mapping across sensory modalities: A neuro-developmental view. *Developmental Neuropsychology*, *41*(5-8), 293–307. https://doi.org/10.1080/87565641.2016.1243112

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, *71*(4), 878–894.

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word-object relations. *Infancy*, *9*(3), 259–288. https://doi.org/10.1207/s15327078in0903_1

Gregory, N. J., Hermens, F., Facey, R., & Hodgson, T. L. (2016). The developmental trajectory of attentional orienting to socio-biological cues. *Experimental Brain Research*, *234*(6), 1351–1362. https://doi.org/10.1007/s00221-016-4627-3

Gregory, S. E. A., & Jackson, M. C. (2017). Joint attention enhances visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(2), 237–249. https://doi.org/10.1037/xlm0000294

Heed, T., Habets, B., Sebanz, N., & Knoblich, G. (2010). Others' actions reduce crossmodal integration in peripersonal space. *Current Biology*, *20*(15), 1345–1349. https://doi.org/10.1016/j.cub.2010.05.068

Heimler, B., van Zoest, W., Baruffaldi, F., Rinaldi, P., Caselli, M. C., & Pavani, F. (2015). Attentional orienting to social and nonsocial cues in early deaf adults. *Journal of Experimental Psychology: Human Perception and Performance, 41*(6), 1758–1771. https://doi.org/10.1037/xhp0000099

Hermens, F. (2017). The effects of social and symbolic cues on visual search: Cue shape trumps biological relevance. *Psihologija, 50*(2), 117–140.

Hobson, P., & Hobson, J. (2011). Joint attention or joint engagement? Insights from autism. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 115–136). Cambridge, MA: MIT Press.

Hoehl, S., Wahl, S., Michel, C., & Striano, T. (2012). Effects of eye gaze cues provided by the caregiver compared to a stranger on infants' object processing. *Developmental Cognitive Neuroscience*, *2*(1), 81–89. https://doi.org/10.1016/J.DCN.2011.07.015

Horstmann, A., & Hoffmann, K.-P. (2005). Target selection in eye–hand coordination: Do we reach to where we look or do we look to where we reach? *Experimental Brain Research*, *167*(2), 187–195. https://doi.org/10.1007/s00221-005-0038-6

Hyde, D. C., Flom, R., & Porter, C. L. (2016). Behavioral and neural foundations of multisensory face-voice perception in infancy. *Developmental Neuropsychology*, *41*(5-8), 273–292. https://doi.org/10.1080/87565641.2016.1255744

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience & Biobehavioral Reviews*, *33*(6), 843–863. https://doi.org/10.1016/j.neubiorev.2009.02.004

Jesse, A., & Johnson, E. K. (2016). Audiovisual alignment of co-speech gestures to speech supports word learning in 2-year-olds. *Journal of Experimental Child Psychology*, *145*, 1–10. https://doi.org/10.1016/j.jecp.2015.12.002

Kaplan, F., & Hafner, V. V. (2006). The challenges of joint attention. *Interaction Studiesinteraction Studies Social Behaviour and Communication in Biological and Artificial Systems*, *7*(2), 135–169.

Keetels, M., & Vroomen, J. (2012). Perception of synchrony between the senses. In M. M. Murray & M. T. Wallace (Eds.), *The neural bases of multisensory processes*. Boca Raton, FL: CRC Press/Taylor & Francis.

Keysers, C., Wicker, B., Gazzola, V., Anton, J.-L., Fogassi, L., & Gallese, V. (2004). A Touching Sight: SII/PV Activation during the Observation and Experience of Touch. *Neuron, 42*(2), 335–346. https://doi.org/10.1016/S0896-6273(04)00156-4

Kim, K., & Mundy, P. (2012). Joint attention, social-cognition, and recognition memory in adults. *Frontiers in Human Neuroscience*, *6*, 172. https://doi.org/10.3389/fnhum.2012.00172

Kinard, J. L., & Watson, L. R. (2015). Joint attention during infancy and early childhood across cultures. In J. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (pp. 844–850). https://doi.org/10.1016/B978-0-08-097086-8.23172-3

Kirkham, N. Z., Rea, M., Osborne, T., White, H., & Mareschal, D. (2019). Do cues from multiple modalities support quicker learning in primary schoolchildren? *Developmental Psychology*, *55*, 2048–2059. https://doi.org/10.1037/dev0000778

Klapetek, A., Ngo, M. K., & Spence, C. (2012). Does crossmodal correspondence modulate the facilitatory effect of auditory cues on visual search? *Attention, Perception, & Psychophysics*, *74*(6), 1154–1167. https://doi.org/10.3758/s13414-012-0317-9

Klein, R. M., & Lawrence, M. A. (2012). On the modes and domains of attention. In M. I. Posner (Ed.), *Cognitive neuroscience of attention,* 2nd (pp. 11–28). New York, NY: Guilford Press.

Kompatsiari, K., Ciardo, F., Tikhanoff, V., Metta, G., & Wykowska, A. (2019). It's in the eyes: The engaging role of eye contact in HRI.

International Journal of Social Robotics, 1–11. https://doi.org/10.1007/s12369-019-00565-4

Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, *4*(2), 50–59. https://doi.org/10.1016/S1364-6613(99)01436-9

Leavens, D., & Racine, T. P. (2009). Joint attention in apes and humans: Are humans unique? *Journal of Conciousness Studies*, *16*(6-8), 240–267.

Liebal, K., Waller, B. M., Burrows, A. M., & Slocombe, K. E. (2014). *Primate communication: A multimodal approach*. https://doi.org/10.1017/CBO9781139018111

Lieberman, A. M., Hatrak, M., & Mayberry, R. I. (2014). Learning to look for language: Development of joint attention in young deaf children. *Language Learning and Development, 10*(1), 19–35. https://doi.org/10.1080/15475441.2012.760381

Mason, G. M., Goldstein, M. H., & Schwade, J. A. (2019). The role of multisensory development in early language learning. *Journal of Experimental Child Psychology, 183*, 48–64. https://doi.org/10.1016/j.jecp.2018.12.011

McDonald, J. J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, *407*(6806), 906–908. https://doi.org/10.1038/35038085

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247–279. https://doi.org/10.1016/0010-0285(82)90010-X

Moore, C., & Dunham, P. J. (1995). Current Themes in Research of Joint Attention. In C. Moore & P. J. Dunham (Eds.), Joint attention: Its origins and role in development (pp. 15–28). Hillsdale, NJ: Lawrence Erlbaum.

Morales, M., Mundy, P., Crowson, M. M., Neal, A. R., & Delgado, C. E. F. (2005). Individual differences in infant attention skills, joint attention, and emotion regulation behaviour. *International Journal of Behavioral Development*, *29*(3), 259–263. https://doi.org/10.1177/01650250444000432

Mundy, P. (2016). *Autism and joint attention: Development, neuroscience, and clinical fundamentals*. Guilford Publications.

Mundy, P. (2018). A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *European Journal of Neuroscience*, *47*(6), 497–514. https://doi.org/10.1111/ejn.13720

Mundy, P., & Jarrold, W. (2010). Infant joint attention, neural networks and social cognition. *Neural Networks*, *23*(8), 985–997. https://doi.org/10.1016/j.neunet.2010.08.009

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science*, *16*(5), 269–274. https://doi.org/10.1111/j.1467-8721.2007.00518.x

Mundy, P., & Sigman, M. (2015). Joint attention, social competence, and developmental psychopathology. In *Developmental psychopathology* (pp. 293–332). https://doi.org/10.1002/9780470939383.ch9

Ngo, M. K., & Spence, C. (2010). Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Attention, Perception, & Psychophysics*, *72*(6), 1654–1665. https://doi.org/10.3758/APP.72.6.1654

Nuku, P., & Bekkering, H. (2010). When one sees what the other hears: Crossmodal attentional modulation for gazed and non-gazed upon auditory targets. *Consciousness and Cognition, 19*(1), 135–143. https://doi.org/10.1016/j.concog.2009.07.012

Núñez, M. (2014). *Joint attention in deafblind children: A multisensory path towards a shared sense of the world*. London: Sense.

O'Madagain, C., Kachel, G., & Strickland, B. (2019). The origin of pointing: Evidence for the touch hypothesis. *Science Advances, 5*(7), eaav2558. https://doi.org/10.1126/sciadv.aav2558

Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science*, *283*(5406), 1272–1273.

Pavani, F., Venturini, M., Baruffaldi, F., Caselli, M. C., & van Zoest, W. (2019). Environmental Learning of Social Cues: Evidence From Enhanced Gaze Cueing in Deaf Children. *Child Development*, *90*(5), 1525–1534. https://doi.org/10.1111/cdev.13284

Pelz, J., Hayhoe, M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, *139*(3), 266–277.

Pfeiffer-Leßmann, N., Pfeiffer, T., & Wachsmuth, I. (2012). An operational model of joint attention: Timing of gaze patterns in interactions between humans and a virtual human. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 851–856). Austin, TX: Cognitive Science Society.

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Prezbindowski, A. K., Adamson, L. B., & Lederberg, A. R. (1998). Joint attention in deaf and hearing 22 month-old children and their hearing mothers. *Journal of Applied Developmental Psychology*, *19*(3), 377–387. https://doi.org/10.1016/S0193-3973(99)80046-X

Richman, A. L., Miller, P. M., & LeVine, R. A. (1992). Cultural and educational variations in maternal responsiveness. *Developmental Psychology, 28*(4), 614–621. https://doi.org/10.1037/0012-1649.28.4.614

Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. *Animal Behaviour*, *58*(5), 921–931. https://doi.org/10.1006/anbe.1999.1242

Scaife, M., & Bruner, J. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*, 265–266.

Schilbach, L. (2015). Eye to eye, face to face and brain to brain: Novel approaches to study the behavioral dynamics and neural mechanisms of social interactions. *Current Opinion in Behavioral Sciences*, *3*, 130–135. https://doi.org/10.1016/J.COBEHA.2015.03.006

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition*, *88*(3), B11–B21. https://doi.org/10.1016/S0010-0277(03)00043-X

Sebanz, N., Knoblich, G., & Prinz, W. (2005). How two share a task: Corepresenting stimulus-response mappings. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1234–1246. https://doi.org/10.1037/0096-1523.31.6.1234

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

Seemann, A. (2019). *The Shared World: Perceptual Common Knowledge, Demonstrative Communication, and Social Space*. Cambridge, MA: MIT Press.

Sellaro, R., Treccani, B., & Cubelli, R. (2018). When task sharing reduces interference: Evidence for division-of-labour in Stroop-like tasks. *Psychological Research*, 1–16. https://doi.org/10.1007/s00426-018-1044-1

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences, 12*(11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shepherd, S. V. (2010). Following gaze: gaze-following behavior as a window into social cognition. *Frontiers in Integrative Neuroscience*, *4*, 5. https://doi.org/10.3389/fnint.2010.00005

Shteynberg, G. (2015). Shared attention. *Perspectives on Psychological Science*, *10*(5), 579–590. https://doi.org/10.1177/1745691615589104

Sinnett, S., Spence, C., & Soto-Faraco, S. (2007). Visual dominance and attention: The Colavita effect revisited. *Perception &*

Psychophysics, *69*(5), 673–686. https://doi.org/10.3758/BF03193770

Siposova, B., & Carpenter, M. (2019). A new look at joint attention and common knowledge. *Cognition*, *189*, 260–274. https://doi.org/10.1016/j.cognition.2019.03.019

Soto-Faraco, S., Sinnett, S., Alsius, A., & Kingstone, A. (2005). Spatial orienting of tactile attention induced by social cues. *Psychonomic Bulletin & Review*, *12*(6), 1024–1031. https://doi.org/10.3758/bf03206438

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and Cognition*, *22*(1), 245–260. https://doi.org/10.1016/j.concog.2012.12.006

Spence, C., McDonald, J., & Driver, J. (2004a). Exogenous spatial-cuing studies of human cross-modal attention and multisensory integration. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 276–320). https://doi.org/10.1093/acprof:oso/9780198524861.003.0011

Spence, C., Pavani, F., Maravita, A., & Holmes, N. (2004b). Multisensory contributions to the 3-D representation of visuotactile peripersonal space in humans: Evidence from the crossmodal congruency task. *Journal of Physiology-Paris*, *98*(1), 171–189. https://doi.org/10.1016/j.jphysparis.2004.03.008

Spencer, P. E. (2000). Looking without listening: Is audition a prerequisite for normal development of visual attention during infancy? *Journal of Deaf Studies and Deaf Education*, *5*(4), 291–302. https://doi.org/10.1093/deafed/5.4.291

Stephenson, L. J., Edwards, S. G., Howard, E. E., & Bayliss, A. P. (2018). Eyes that bind us: Gaze leading induces an implicit sense of agency. *Cognition, 172*, 124–133. https://doi.org/10.1016/j.cognition.2017.12.011

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology*, *55*(1), 96–109. https://doi.org/10.1037/dev0000628

Swingler, M. M., Perry, N. B., & Calkins, S. D. (2015). Neural plasticity and the development of attention: Intrinsic and extrinsic influences. *Development and Psychopathology*, *27*(2), 443–457. https://doi.org/10.1017/S0954579415000085

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences, 14*(9), 400. https://doi.org/10.1016/J.TICS.2010.06.008

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Tomasello, M. (2014). *A natural history of human thinking*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. Cambridge, MA: Harvard University Press.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*(6), 1454–1463. https://doi.org/10.2307/1130423

Tschentscher, N., & Fischer, M. H. (2008). Grasp cueing and joint attention. *Experimental Brain Research*, *190*(4), 493–498. https://doi.org/10.1007/s00221-008-1538-y

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1053–1065. https://doi.org/10.1037/0096-1523.34.5.1053

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2009). Poke and pop: Tactile–visual synchrony increases visual saliency. *Neuroscience Letters*, *450*(1), 60–64. https://doi.org/10.1016/j.neulet.2008.11.002

Van der Burg, E., Cass, J., Olivers, C. N. L., Theeuwes, J., & Alais, D. (2010). Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PLoS ONE*, *5*(5), e10664. https://doi.org/10.1371/journal.pone.0010664

Stoep, N. van der, Postma, A., & Nijboer, T. C. W. (2017). Multisensory perception and the coding of space. In A. Postma & I. van der Ham (Eds.), *Neuropsychology of space: Spatial functions of the human brain* (pp. 123–158). https://doi.org/10.1016/B978-0-12-801638-1.00004-5

Volpe, G., & Gori, M. (2019). Multisensory interactive technologies for primary education: From science to technology. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.01076

Vroomen, J., & De Gelder, B. (2004). Perceptual effects of cross-modal stimulation: Ventriloquism and the freezing phenomenon. In *The handbook of multisensory processes* (Vol. 3, pp. 1–23). The MIT Press, London.

Vygotsky, L. S. (2012). *Thought and language* (E. Hanfmann, G. Vakar, & A. Kozulin, Eds.). Cambridge, MA: MIT Press.

Wahl, S., Marinović, V., & Träuble, B. (2019). Gaze cues of isolated eyes facilitate the encoding and further processing of objects in 4-month-old infants. *Developmental Cognitive Neuroscience*, *36*, 100621. https://doi.org/10.1016/j.dcn.2019.100621

Wahn, B., & König, P. (2017). Can limitations of visuospatial attention be circumvented? A review. *Frontiers in Psychology*, *8*, 1896. https://doi.org/10.3389/fpsyg.2017.01896

Wahn, B., Keshava, A., Sinnett, S., Kingstone, A., & König, P. (2017). Audiovisual integration is affected by performing a task jointly. *Proceedings of the 39th annual conference of the cognitive science society*, 1296–1301.

Willemse, C., Marchesi, S., & Wykowska, A. (2018). Robot faces that follow gaze facilitate attentional engagement and increase their likeability. *Frontiers in Psychology*, *9*, 70. https://doi.org/10.3389/fpsyg.2018.00070

Wright, R. D., & Ward, L. M. (2008). *Orienting of attention*. Oxford: Oxford University Press.

Yang, G.-Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., … Wood, R. (2018). The grand challenges of Science Robotics. *Science Robotics*, *3*(14), eaar7650. https://doi.org/10.1126/scirobotics.aar7650

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS ONE*, *8*(11), e79659. https://doi.org/10.1371/journal.pone.0079659

Yu, C., Schermerhorn, P., & Scheutz, M. (2012). Adaptive eye gaze patterns in interactions with human and artificial agents. *ACM Transactions on Interactive Intelligent Systems*, *1*(2), 1–25. https://doi.org/10.1145/2070719.2070726

Chapter 5

Paper IV: The Impact of Joint Attention on the Sound-Induced Flash Illusions

Battich, L., Garzorz, I., Wahn, B., and Deroy, O. (forthcoming 2021). The impact of joint attention on the sound-induced flash illusions. *Attention, Perception and Psychophysics*.

Author contributions:

L.B., B.W. and O.D. conceived of the research idea. L.B., I.G. and O.D. designed the study. L.B. carried out the experiments, analysed the data, and wrote the paper with help from all authors.

The Impact of Joint Attention on the Sound-Induced Flash Illusions

Lucas Battich[1, 2], Isabelle Garzorz[2], Basil Wahn[3, 4], Ophelia Deroy[2, 5, 6]

[1]Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität München

[2]Faculty of Philosophy & Philosophy of Science, Ludwig-Maximilians-Universität München

[3]Department of Psychology, University of British Columbia

[4]Department of Psychology, Leibniz Universität Hannover

[5]Munich Center for Neuroscience, Ludwig-Maximilians-Universität München

[6]Institute of Philosophy, School of Advanced Studies, University of London

Author Note

We have no known conflicts of interest to disclose.

We acknowledge the support of a DFG research fellowship (WA 4153/2-1) awarded to BW. OD was supported by a grant from the Excellence Initiative in the LMU, and a grant from the NOMIS foundation (acronym DISE).

Correspondence concerning this article should be addressed to Lucas Battich, Faculty of Philosophy & Philosophy of Science, LMU Munich, Geschwister-Scholl-Platz 1, 80359 München, Germany. Telephone: +49 89-2180-3282. Email: Lucas.Battich@campus.lmu.de

Abstract

Humans continuously coordinate where they focus their attention in the environment with others. In many cases, as when listening to a concert, or hunting together, they also need to select and integrate information from different senses, or ignore one modality to focus on a task-relevant one. Here we examine how joint attention modulates multisensory integration. In this preregistered study, we test whether the prevalent hypothesis that joint attention enhances stimulus information encoding and processing, over and above individual attention, extends to temporal multisensory integration. We used the sound-induced flash illusions, where an incongruent number of visual flashes and auditory beeps induces a single flash to be seen as two (fission illusion), and two flashes as one (fusion illusion). By asking participants to count flashes either alone or together, we expected that enhanced processing of the visual target relative to the distracting accompanying sounds would lead to a decrease of both fission and fusion illusions when the targets were jointly attended. We found that joint attention did not affect the overall frequency of illusions, but decreased participants' criterion bias in the fusion illusion. Our results reveal the limitations of the theory that joint attention results in greater processing resources as it does not extend to temporal audiovisual integration.

Keywords: joint attention, multisensory integration, sound-induced flash illusion

The Impact of Joint Attention on the Sound-Induced Flash Illusions

1. Introduction

People devote greater cognitive resources to those features in their environment that are co-attended simultaneously with others (Becchio, Bertone, & Castiello, 2008; Shteynberg, 2015, 2018). Known as joint attention, coordinating attention with others on a common target, even in the absence of communication, enhances a participant's mental spatial rotation performance (Böckler, Knoblich, & Sebanz, 2011), and facilitates information encoding in working memory (Gregory & Jackson, 2017; Kim & Mundy, 2012). The prevalent theoretical hypothesis regarding the functional role of joint attention is therefore that it deepens or enhances the encoding of stimulus information in ways that are not observed when information is individually attended (Mundy, 2016, 2018; see also Becchio et al., 2008; Shteynberg, 2015). This hypothesis explains why joint attention plays a fundamental role in language acquisition, the development of theory of mind, and the ability to engage in more complex activities with others (Bottema-Beutel, 2016; Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998; Mundy & Newell, 2007).

The hypothesis of an 'encoding enhancement' also accords with findings on the influence of gaze-based joint attention on perceptual judgements. Gaze-cueing studies using covert orienting paradigms show that another's gaze behaviour can influence detection and discrimination of visual stimuli (see Frischen, Bayliss, & Tipper, 2007 for a review). For example, when the subject's and an avatar's spatial visual attention is in conflict, response times to judge the number of stimuli presented are slower (Samson et al., 2010). While most studies use response times as their primary dependent measure, Seow & Fleming (2019) report that participants' perceptual sensitivity (*d'*) for detecting Gabor patches increased when it was co-witnessed with a bystander agent with a congruent perspective.

In everyday situations, however, many joint attention scenarios also involve multisensory targets of attention, where information from different senses has either to be selected and integrated or, on the contrary, separated (Battich, Fairhurst, & Deroy, 2020). Previous work addressing the multisensory aspects of joint attention in adults has focused predominantly on spatial judgments. Using computer avatar's eye-gaze cues, Soto-Faraco, Sinnett, Alsius, & Kingstone (2005) have shown that both detection and discrimination of tactile stimuli at the body location congruent with the other's gaze direction is facilitated over the incongruent body location. Extending these results, Nuku & Bekkering (2010) show that task-irrelevant directional gaze cues from a virtual partner influence perceptual judgement also in the auditory modality, such that processing stimuli at the jointly attended location is facilitated.

Taken together, these findings accord with the hypothesis that joint attention results in a socially-mediated enhancement of the relative encoding and processing of co-attended sensory stimuli, compared to solo attention (Mundy, 2016; 2018). The precise extent of this hypothesis for multisensory processing remains largely untested. In this study, we examine whether this hypothesis extends to temporal multisensory processing. Specifically, we examined whether jointly attending to the visual component of multisensory events would also result in enhancing its processing, or would reduce the weight of jointly presented, but not jointly attended sounds. Though target enhancement and reduction of distractors are often considered two sides of the same coin, mechanistic differences provide reasons to regard them as possibly distinct phenomena (Chelazzi, Marini, Pascucci, & Turatto, 2019; Noonan et al., 2016; van Moorselaar & Slagter, 2020).

Joint attention is often operationalized in terms of visually perceiving where another person (or an avatar) is gazing. Tracking where someone is attending is nonetheless neither necessary nor sufficient for joint attention, which involves representing that the two co-

attenders attend to the same perceptual target (Carpenter et al., 1998; Mundy, 2018; Siposova & Carpenter, 2019; Tomasello, 1995). It is possible, for instance, that one monitors someone else's gaze without them noticing, which is then not a case of joint attention. It is also possible that joint attention occurs when both agents realize that they are attending to the same object, even though they are not closely monitoring each other's gaze. In other words, joint attention has more to do with the representation of a "social locus of attention" than with gaze-following, even in the absence of verbal communication. To extend previous research based on artificial avatars' shared gaze, it is necessary to investigate how this shared locus of attention between two people affects the processing of multisensory information. In the present study, we operationalize joint attention as the situation in which two individuals focus their perceptual attention on the same modal target, and both know that they are attending the same target.

Relatedly, previous studies have shown that multisensory processing in visuotactile (Heed, Habets, Sebanz, & Knoblich, 2010) and audiovisual (Wahn, Keshava, Sinnett, Kingstone, & König, 2017) spatial interference tasks is affected by a division of labour manipulation. In these spatial tasks, stimuli in different sensory modalities were simultaneously presented either in congruent or incongruent locations. For incongruent presentations, visual stimuli tended to distract/interfere with tactile and auditory localisation judgements as humans generally tend to rely more on visual information for spatial tasks. In a collective condition, each participant in a pair had to attend and respond to a target stimulus in different modalities, while ignoring the other modality. That is, one person would be tasked to locate auditory stimuli (Wahn et al., 2017) or tactile stimuli (Heed et al., 2010) while the other person would be tasked to locate visual stimuli. When participants performed their respective tasks together, participants performing the tactile (Heed et al., 2010) and

auditory (Wahn et al., 2017) localisation task were better able to ignore the visual distractor stimuli compared to a condition when they performed the same task on their own.

A recent study extended this division of attentional labour to the sound-induced flash illusion task (Wahn, Rohe, Gearhart, Kingstone, & Sinnett, 2020), which has two variants: fission, where a single flash accompanied by two auditory beeps induces a visual percept of the flashes, and fusion, where two flashes are perceived as one when accompanied by one beep (Andersen, Tiippana, & Sams, 2004; Shams, Kamitani, & Shimojo, 2000). The rationale behind the illusions is that the auditory signal dominates over the visual signal in tasks requiring temporal precision, altering the integrated percept (Andersen et al., 2004; Shams, Kamitani, & Shimojo, 2002).

As for the studies mentioned above (Heed et al., 2008; Wahn et al., 2017), tasks were divided along sensory modalities. That is, one participant was asked to count the number of flashes presented, while a confederate was required to count the number of beeps. When these tasks were performed together, participants perceived the sound-induced fission illusion significantly more often compared to performing the flash counting task alone. However, this effect was no longer found when a divider was placed between the participant and confederate, suggesting that visual access is critical. Taken together with the studies mentioned above (Heed et al., 2010; Wahn et al., 2017) where participants were better able to ignore distracting *visual* stimuli, the authors suggest that the presence of another person may act as a visual distractor such that visual information (presented on the computer screen) is attended to a lesser extent. Depending on the performed task, this can either lead to less distraction by visual distractors when participants performed an auditory or tactile localisation task (Heed et al., 2010; Wahn et al., 2017) or more distraction by auditory stimuli when participants performed a visual flash counting task (Wahn et al., 2020), leading to an increase in the perceived fission illusions. There is no previous indication, to our knowledge,

of whether a similar or opposite effect would be observed in a joint attention manipulation, where both participants in a pair are required to attend and respond to the same modal target.

Here, we address whether the hypothesis that joint attention can boost relative processing of sensory of co-attended stimuli (by facilitating the processing of the jointly attended modality and/or reducing the distraction to the non-attended modality) compared to solo attention applies to temporal multisensory processing. In this preregistered study (preregistration available at https://osf.io/v5gjp), we investigate this question using the sound-induced flash illusions. Not only does this make the comparison with the division of attentional labour possible, but sound-induced flash illusions are also reliable indicators of the multisensory integration of temporally aligned stimuli (Keil, 2020; Hirst et al., 2020). What is more, the fact that the illusion is based on time, rather than space, avoids colluding attention to a given modality and attention to a distinct region of space.

While the specific interactions between attention and multisensory processes are a matter of ongoing debate, mounting evidence suggests that multisensory integration can be modulated by attentional control (for reviews, see, Choi, Lee, & Lee, 2018; Macaluso et al., 2016; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010). Several studies report possible cognitive influences on the sound-induced flash illusions (for reviews see Keil, 2020; Hirst et al. 2020), yet earlier studies only investigated manipulations of the individual's attentional focus. For instance, Andersen et al. (2004) found that the integration of audiovisual information during both fission (one flash seen as two) and fusion (two flashes seen as one) illusions was not automatic but varied depending on whether participants were asked to count beeps or flashes. They thus suggest that the illusions are susceptible to differences in attentional control, an interpretation supported by findings that the fission illusion is modulated by selective spatial attention (Mishra, Martínez, & Hillyard, 2010; Odegaard, Wozny, & Shams, 2016). The fission illusion is also modulated by cognitive load (Michail &

Keil, 2018) and top-down expectations about the proportion of illusion-inducing trials (Wang et al., 2019).

The present study aims to extend our understanding of how attentional and social factors affect the fusion and fission illusions by using a joint attention manipulation. As current functional models of joint attention suggest that sharing the locus of attention with another person will enhance information processing in ways that solo attention does not (Mundy, 2018; Battich et al. 2020), it is important to investigate to what extent a joint attention manipulation may affect multisensory processing. Investigating such manipulation is relevant not only for experimenters as they may need to reconsider the possible effects of being within a participant's view while testing the sound-induced flash illusions (Hirst et al., 2020), but also for daily life as we often perceive multisensory stimuli in social situations. If engaging in joint attention enhances processing of a jointly attended visual target (Becchio et al., 2008; Mundy, 2016, 2018; Shteynberg, 2015, 2018), we predict a shift in the relative weighting of visual and auditory information, so that the strength of the sound-induced illusions will be reduced during joint attention compared to performing the task alone. In accordance with maximum-likelihood-estimation (Ernst & Bülthoff, 2004) and Bayesian inference (Shams & Kim, 2010) frameworks, this shift in audio-visual integration could either lead to a boost in processing of the visual target and/or reduce processing of the auditory distractor (van Moorselaar & Slagter, 2020). As a control condition, we also expect that the mere co-presence of another person who is not engaged in joint attention with the participant, will not affect the illusions compared to individual performance.

Finally, we had no specific predictions regarding possible differences between fission and fusion illusions. Known neural (Mishra, Martinez, & Hillyard, 2008) and behavioural differences between the two illusion variants suggest that we should not treat them as necessarily identical. Neuroimaging studies show that the fission illusion correlates with

activity modulation in early visual cortical areas and superior colliculus, suggesting that the illusion underlies a multisensory process involving early perceptual stages (Cecere, Rees, & Romei, 2015; Shams, Iwaki, Chawla, & Bhattacharya, 2005; Watkins, Shams, Tanaka, Haynes, & Rees, 2006; Zhang & Chen, 2006). The fusion illusion correlates with activity in retinotopic primary visual cortex and the superior temporal sulcus (Watkins, Shams, Josephs, & Rees, 2007), a brain region associated with multisensory integration (Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004). Susceptibility to the fission illusion, but not the fusion illusion, varies with age (DeLoss & Andersen, 2015; McGovern, Roudaia, Stapleton, McGinnity, & Newell, 2014) and with emotionally-charged stimuli (Takeshima, 2020). Importantly, there is preliminary evidence that cognitive expectations (Wang et al., 2019) decrease the occurrence of fission but not fusion illusions. Given these potential differences in the mechanisms underlying the fission and fusion illusions, our study will test the influence of joint attention for both illusions.

2. Material and methods

2.1. Participants

Given current literature on possible social effects on the sound-induced flash illusion, our estimate of a Cohen's *d* effect size is of around 0.41 (Wahn et al., 2020). We used the software G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) to conduct a power analysis, to obtain .80 power to detect Cohen's *d* effect size of 0.415 for a two-tailed paired t-test, at the standard .05 alpha error probability. Based on this, our target sample size was forty-eight participants. Due to the possibility of some participants not meeting the inclusion criteria, we recruited fifty-two volunteers (29 female, 1 undisclosed gender, M = 27.96 years, SD = 5.9 years) to participate in the study. Participants received either 9 EUR or course credits as compensation for their participation, at their choice. All participants had normal or corrected-to-normal vision and hearing, and were right-handed, with mean handedness score M =

95.26, SD = 15.18, as measured by the shortened Edinburgh Handedness Inventory (Oldfield, 1971; Veale, 2014).

The study was conducted in accordance with the Declaration of Helsinki and approved by the ethics committee of the University of London (approval ref. SASREC_1819_313A). All participants gave written informed consent before their participation.

2.2. Materials

Pairs of participants sat next to each other in front of the same computer screen, (model Asus VG248QE 24 inches, of 1920 x 1080 pixels resolution, and 60 Hz refresh rate), and at a fixed viewing distance (60 cm) from the screen. Their heads were aligned to the outer edges of the screen (width 53 cm), so that when looking straight ahead they see the screen outer edge. Two speakers (model Logitech Z200) were set adjacent to each side of the screen so that the speaker's middle was levelled with the lower edge of the screen.

A fixation cross was presented for an interval that varied randomly between 1000-1400 ms, followed by the visual and auditory stimuli. The visual stimulus consisted of a uniform white disc (radius of 2° of visual field, positioned 5° below the fixation cross), flashed for 17 ms, on a black computer screen. The auditory stimulus consisted of a sine-wave beep of 7 ms duration with 3.5 kHz frequency. Stimulus onset asynchrony (SOA) for consecutive stimuli was 57 ms for sound beeps, and 67 ms for visual flashes. The first beep was presented always 23 ms prior to the first flash (Figure 1).

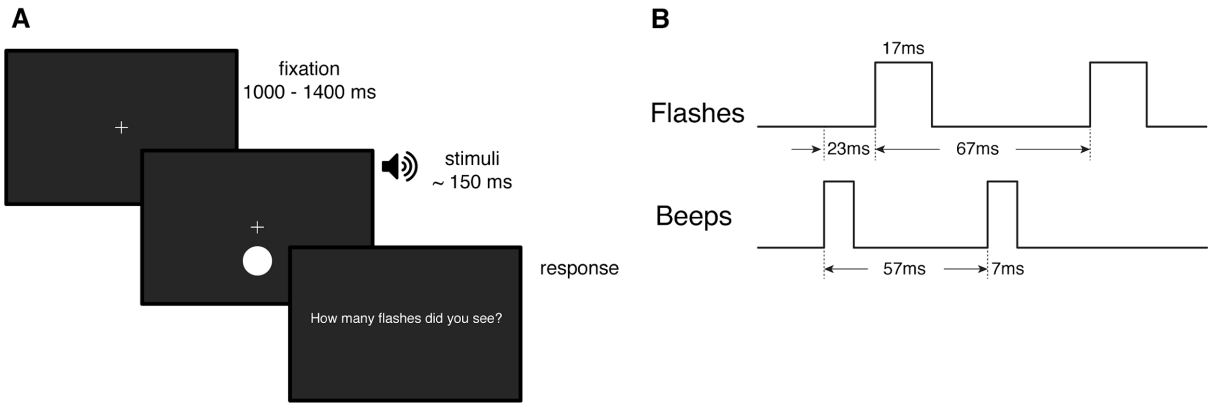

Figure 1. (A) Single trial procedure and (B) temporal order of stimuli (when two flashes and two beeps are presented).

2.3. Procedure

In each trial, either 1 or 2 flashes were presented, accompanied by either 1 or 2 beeps, giving 4 types of trials (1F1B, 1F2B, 2F1B, 2F2B). Each of the 4 types of trials was presented 30 times. The 120 trials were fully randomized and presented in 4 blocks with approx. 10 seconds rest between blocks. Participants were asked to judge how many visual flashes they saw, by clicking the left or right buttons of a computer mouse allocated to each participant, to report one or two flashes, respectively. Both participants were given the same instructions simultaneously and knew that they were performing the same task.

Participants performed the full set of 120 trials three times, one per social condition: individually, jointly, and during a co-presence control. In the individual condition, participants sat alone to perform the task, in the same seat that they occupy during the joint attention and co-presence control conditions (i.e., a given participant always had the same seat); the second participant waited in a separate testing room. During the joint attention condition, both participants were instructed to attend to the visual stimuli and perform the task concurrently. Each participant still provided their answer individually. In the co-presence control condition, participants sat side by side as in the joint attention condition but oriented in opposite directions. One participant performed the flash-counting task, while the second

participant performed an unrelated drawing task on paper. Then participants switched roles. During the 120 trials of each condition, the experimenter waited outside the testing room, out of sight from both participants. Participants were instructed to avoid talking to each other during the flash-counting task. Due to the fast duration of each trial and the demanding nature of the flash-counting task, verbal communication is also very difficult to achieve. After 120 trials were completed, participants saw a text on the screen requesting that the experimenter should be contacted. The experimenter then made the necessary setup adjustment depending on the next social condition, instructing each participant on their assigned role (e.g. to perform the flash-counting task, wait in an adjacent room, or perform an unrelated drawing task). The order of social conditions was counterbalanced across participants. In most cases, the session took approximately 45 minutes. The experiment was programmed using Python (version 3.6.8) and the PsychoPy library (version 3.2.3; Peirce, 2007; Peirce et al., 2019).

2.4. Data analysis plan

To analyze the effect of shared attention on the strength of the illusions, we preregistered to conduct a 2x3 repeated measures ANOVA for the mean responses with beeps (1, 2 beeps) and social condition (individual, joint attention, control) as within-subject factors, separately for the fission (1 flash trials) and fusion (2 flashes trials) illusions.

We also pre-registered and planned two paired t-test comparisons over the interaction effects between beeps and social conditions on the number of flashes perceived. First, to test whether joint attention reduced the illusions, we contrasted the effect of beeps on the number of flashes reported across the individual and joint attention condition. Second, to test whether the mere presence of another participant affects the frequency of the illusions, we contrasted the effect of beeps on the number of flashes reported across the individual and control condition. We performed these planned comparisons regardless of whether the omnibus interaction was significant (Abelson & Prentice, 1997; Schad et al. 2020).

Since the assumption of normality in the parametric models for the number of flashes perceived (ANOVAs and t-tests) was violated (Shapiro-Wilk tests performed on the data were significant, all $ps < .001$), we conducted permutation-based ANOVAs separately for each illusion (1 flash and 2 flashes trials). We then performed the two planned pairwise comparisons with permutation-based t-tests, for each illusion. Though all comparisons were planned, we report *p*-values corrected using the Bonferroni correction.

We excluded three participants from the sample due to low performance (greater than or equal to 35% incorrect responses) on either or both of congruent trials combinations (equal number of flashes and beeps presented), aggregated across social conditions, probably due to lack of motivation or task compliance. Only single trials with reaction times between 100 ms and 3000 ms were included in the analyses. We thus excluded 0.5% of trials (92 trials) spread over 24 participants from further analyses.

To follow upon performance analyses, we preregistered to conduct exploratory analyses on any possible effects over reaction times across the different experimental manipulations. To examine performance measures that better account for possible dissociations in sensitivity and criterion biases, we also preregistered to analyze possible effects on signal detection measures in the ability to discriminate between one and two flashes, during two-beep trials (coded as Fission), and one-beep trials (coded as Fusion).

Witt and colleagues (Witt, Taylor, Sugovic, & Wixted, 2015, 2016) suggest that the SIFIs should be reflected primarily in the criterion measure as indicative of perceptual processes. Theoretically, the number of beeps biases visual perception to detect the same number of flashes, rather than making visual perception less sensitive per se. Knotts & Shams (2016) suggest that both *d'* and c may reflect perceptual aspects associated with the illusion. An analysis of sensitivity and criterion can therefore provide nuanced measures for testing the impact of social conditions on the illusions. It would be a mistake, however, to interpret

the criterion bias as a decision bias, response-based bias, or a memory bias. Witt et al. (2015; 2016) show that the sound-induced flash illusions are predominantly manifested in the criterion measure *c*, but we are not able to distinguish by SDT techniques alone if this bias is perceptual or decisional. Apart from taking into account these considerations for the interpretation of our results, we also adopt Witt and colleagues' suggestion to compare one-beep versus two-beeps trials (Witt et al., 2016), as we are primarily interested in differences in multisensory processing between social conditions. The single flash stimulus was treated as the target, so that a correct response of one flash when one flash was presented was counted as a hit, and an incorrect response of one flash when two flashes were presented was counted as a false alarm. Sensitivity was defined as $d' = z(H) - z(FA)$, and criterion bias was defined as $c = -.5(z(H) + z(FA))$, where z is the inverse of the cumulative normal. Hit and false alarm rates of 0 and 1 were corrected to $(2N)^{-1}$ and $1 - (2N)^{-1}$, respectively, where N is the number of trials on which the rate is based (Macmillan & Creelman, 2005).

For each illusion, we performed one-way repeated measures ANOVAs of *d'* and *c*, dependent on social condition as a within-subject factor. As in our performance analysis, we then conducted two planned pairwise comparisons (individual vs. joint attention, and individual vs. control), reported with Bonferroni corrected *p*-values.

3. Results

3.1. Fission illusion

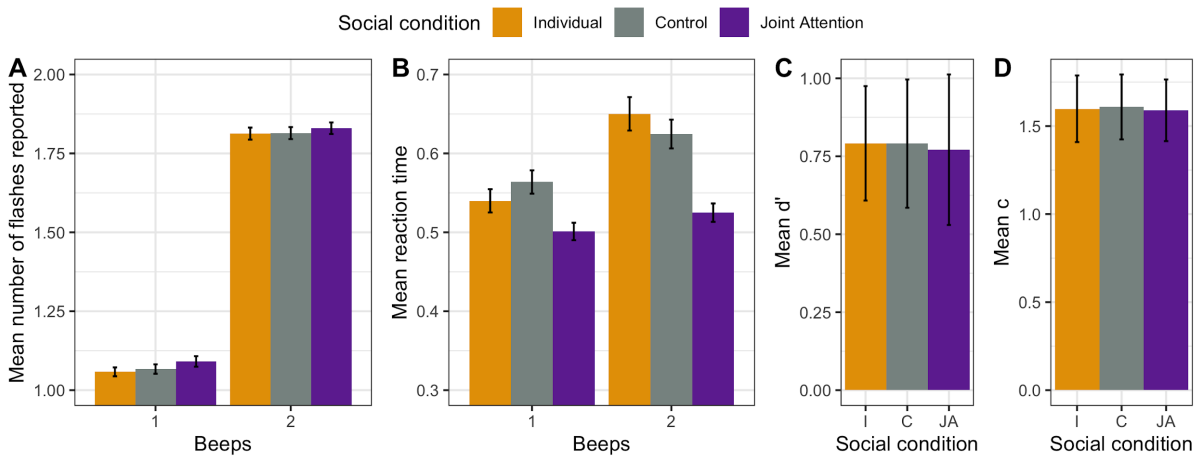

Figure 2. Fission illusion results. (A) Mean number of flashes reported and (B) mean reaction times in 1F1B and 1F2B trials across conditions, in seconds. (C,D) Signal detection measures of sensitivity (*d'*) and bias (*c*) in the ability to discriminate between one and two flashes during 1F2B and 2F2B trials. Error bars show within-subjects adjusted 95% confidence intervals (Cousineau, 2005; Morey, 2008).

Number of flashes perceived

Fig. 2A shows the overall mean of each participant's mean responses in trials where a single flash was presented. Trials with two beeps display a strong increase in the average number of flashes reported. Table 1 shows the mean number of flashes reported and reaction times for all conditions. To test the effect of the social manipulations, we subjected the number of flashes perceived in 1-flash trials to a permutation-based repeated measures ANOVA (Kherad-Pajouh & Renaud, 2015) with beeps (1, 2 beeps) and social condition (individual, joint attention, control) as within-subject factors. We found a significant main effect of beeps, $F(1, 48) = 521.78$, $p < .001$, $\eta_g^2 = .8$. When one flash was presented, the number of beeps affected the number of flashes reported, showing that this audiovisual

manipulation successfully induced a fission illusion. However, we did not find a significant main effect of social condition ($F(2, 96) = 2.41$, $p = .09$, $\eta_g^2 = .004$), nor an interaction effect ($F(2, 96) = 0.16$, $p = .84$, $\eta_g^2 < .001$).

Although the interaction was not significant, we performed the preregistered planned permutation-based paired t-test on the effect of beeps on the number of flashes reported (the difference in responses across one and two beeps trials) between the individual and joint attention conditions. Contrary to our hypothesis, we found no significant difference, $t(48) = -0.45$, *corrected* $p = 1$, Cohen's $d = 0.06$. As these results suggest that engaging in joint attention does not affect susceptibility to the fission illusion, we also computed Bayes factors (BF) for this effect to assess relative likelihoods of the null (H0) and alternative (H1) hypotheses (we note that Bayes factor analyses were not included in our preregistration). BF = 1 indicates equal support for H1 and H0, while BFs between 1-3, 3-10 and > 10 indicate anecdotal, moderate and strong support for H1 respectively, and BFs between .33–1, .1–.33 and < .1 indicate anecdotal, moderate and strong support for H0, respectively (Aczel, Palfi, & Szaszi, 2017). We found a Bayes factor of .17, indicating that our data gives moderate support for the null hypothesis (it is 5.88 more likely under the null than under the alternative hypothesis).

As expected, we found no significant differences in the pairwise comparisons between individual and control conditions on the difference in responses across one and two beeps trials, $t(48) = -0.20$, *corrected* $p = 1$, Cohen's $d = 0.03$. A computed Bayes factor of .16 indicates moderate support for the null hypothesis, so that our data are 6.3 times more likely under the null than under the alternative hypothesis. These results suggest that participants were susceptible to the fission illusion, but this susceptibility did not differ between social conditions.

Table 1

Mean Number of Flashes Reported and Mean Response Times (RTs) for Each Stimulus Type Across Social Conditions.

	Individual		Control		Joint attention	
Stimulus	Flashes reported	RTs (sec.)	Flashes reported	RTs (sec.)	Flashes reported	RTs (sec.)
1F1B	1.06 (0.23)	0.54 (0.29)	1.07 (0.25)	0.56 (0.29)	1.09 (0.29)	0.5 (0.2)
1F2B	1.81 (0.39)	0.65 (0.44)	1.81 (0.39)	0.62 (0.38)	1.83 (0.38)	0.52 (0.22)
2F1B	1.38 (0.48)	0.65 (0.4)	1.39 (0.49)	0.65 (0.4)	1.41 (0.49)	0.54 (0.21)
2F2B	1.96 (0.19)	0.58 (0.37)	1.95 (0.21)	0.55 (0.33)	1.95 (0.21)	0.49 (0.2)

Note. Standard deviations are included in parentheses.

Reaction times

Fig. 2B shows the overall mean of each participant's mean reaction times in trials where a single flash was presented. To test whether the observed difference in latencies across social conditions was significant, we subjected the reaction times to a permutation-based repeated measures ANOVA with beeps (1, 2 beeps) and social condition (individual, joint attention, control) as within-subject factors. We found a significant main effect of beeps

($F(1, 48) = 9.69$, $p < .01$, $\eta_g^2 = .03$), and a significant effect of social condition ($F(2, 96) = 10.45$, $p < .001$, $\eta_g^2 = .04$). The interaction effect was small though significant with $F(2, 96) = 7.78$, $p < .001$, $\eta_g^2 = .009$. We followed this interaction effect with three permutation-based pairwise comparisons, comparing the difference between congruent (1 flash, 1 beep) and incongruent (1 flash, 2 beeps) presentations between social conditions. We found that this congruent-incongruent difference was significantly reduced in the joint attention condition compared to the individual condition ($t(48) = -3.43$, *corrected* $p = .006$, Cohen's $d = 0.48$), and did not significantly differ between individual and control conditions, ($t(48) = -2.27$, *corrected* $p = .078$, Cohen's $d = 0.32$), nor between control and joint attention conditions, ($t(48) = -2.01$, *corrected* $p = .19$, Cohen's $d = 0.28$). Our results indicate that, for comparable performance, the response speed difference between congruent and incongruent trials observed during individual condition disappeared in the joint attention condition.

Signal detection measures

Signal detection theory analysis indicated that sensitivity and criterion bias did not visibly differ across social conditions (Fig. 2, panels C and D, respectively). One-way repeated ANOVAs showed neither a significant effect of social condition on sensitivity *d'* ($F(2,96) = 0.03$, $p = .96$) nor on criterion *c* ($F(2,96) = 0.03$, $p = .96$). Similarly, our pre-planned pairwise comparisons did not reveal significant differences for *d'* and *c* (Table 2).

Table 2

Pairwise Compassions of Signal Detection Measures Across Social Conditions for the Fission Illusion

Measure	Comparison	*t*	*df*	95% CI	Cohen's *d*	Corrected *p*
Sensitivity *d'*	Individual vs joint attention	0.22	48	[-0.17, 0.21]	0.03	1
	Individual vs control	0.01	48	[-0.16, 0.16]	0.00	1
Criterion *c*	Individual vs joint attention	0.11	48	[-0.15, 0.17]	0.02	1
	Individual vs control	-0.13	48	[-0.17, 0.15]	0.02	1

Note. CI = confidence interval; Bonferroni corrected *p* values.

3.2. Fusion illusion

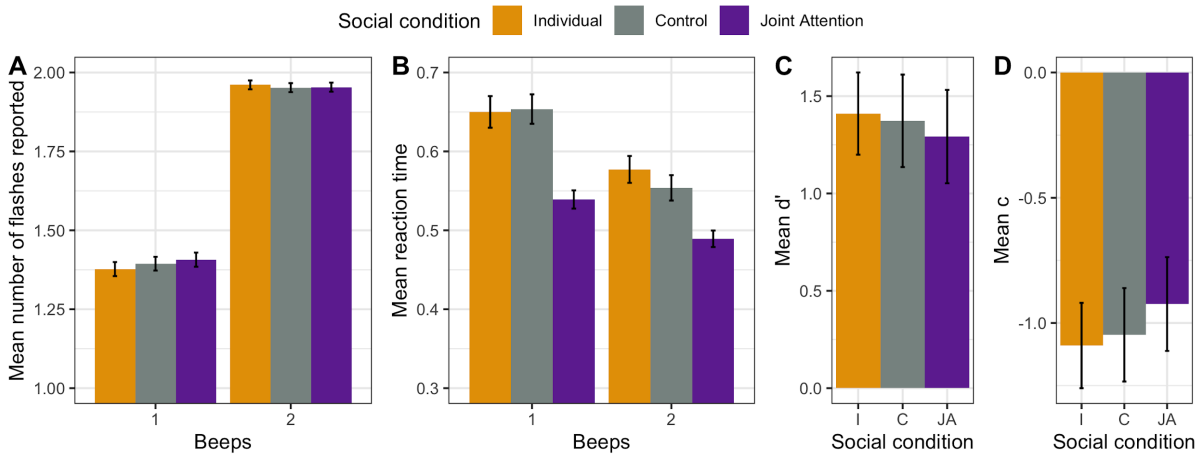

Figure 3. Fusion illusion results. (A) Mean number of flashes reported and (B) mean reaction

times in 2F1B and 2F2B trials across conditions. (C,D) Signal detection measures of sensitivity (*d'*) and bias (*c*) in the ability to discriminate between one and two flashes during 1F2B and 2F2B trials. Error bars show within-subjects adjusted 95% confidence intervals (Cousineau, 2005; Morey, 2008).

Number of flashes perceived

We performed the same analyses to assess the effect on the fusion illusion, where two flashes were presented, as we did for the fission illusion. Fig. 3A shows the overall mean of each participant's mean responses in two-flashes trials. Trials with one beep showed a decrease in the average number of flashes reported, so that when two flashes and one beep were presented concurrently, participants tended toward reporting one flash. Table 1 shows the mean number of flashes reported and reaction times for all conditions. We subjected the number of flashes reported in two-flashes trials to a permutation-based repeated measures ANOVA with beeps (1, 2 beeps) and social condition (individual, joint attention, control) as within-subject factors. We found a significant main effect of beeps on the mean flashes reported, $F(1, 48) = 144.59$, $p < .001$, $\eta_g^2 = .55$, showing that participants were susceptible to the fusion illusion. However, we did not find a significant main effect of social condition ($F(2, 96) = 0.26$, $p = .78$, $\eta_g^2 < .001$) or interaction effect ($F(2, 96) = 0.9$, $p = .4$, $\eta_g^2 < .001$).

Although the interaction effect was not significant, we performed two planned comparisons as pre-registered. To test the hypothesis that the illusion diminishes during joint attention as compared to individual performance, we computed the difference in responses across one and two beeps trials, and then performed a permutation-based paired t-test between the individual and joint attention conditions. Contrary to our hypothesis, we found no significant differences, $t(48) = 1.49$, *corrected* $p = .22$, Cohen's $d = 0.21$. These results suggest that engaging in joint attention does not affect susceptibility to the fusion illusion. To

assess the relative likelihoods of the null and alternative hypotheses, we computed Bayes factors for this comparison and found that with a Bayes factor of .43, our data provides anecdotal support in favour of the null hypothesis, so that the data is 2.27 more likely under the null than the alternative hypothesis. As expected, we found no significant differences in the pairwise comparisons between individual and control conditions on the difference in responses across one and two beeps trials, $t(48) = -0.83$, *corrected* $p = .85$, Cohen's $d = 0.12$. A computed Bayes factor of .21 indicates moderate support for the null hypothesis, so that our data are 4.64 times more likely under the null than the alternative hypothesis.

These results suggest that participants were susceptible to the fusion illusion, yet this susceptibility did not differ between social conditions. Unlike the results for the fission illusion, however, our data provides only anecdotal support for the null hypothesis that there are no differences between individual and joint attention conditions.

Reaction times

Fig. 3B shows the overall mean of each participant's mean reaction times in trials where two flashes were presented. We subjected the reaction times to a permutation-based repeated measures ANOVA with beeps (1, 2 beeps) and social condition (individual, joint attention, control) as within-subject factors. We found a significant main effect of beeps ($F(1, 48) = 32.64$, $p < .001$, $\eta_g^2 = .03$), and a significant effect of social condition ($F(2, 96) = 11.62$, $p < .001$, $\eta_g^2 = .05$); yet the interaction effect was not significant ($F(2, 96) = 2.23$, $p = .11$, $\eta_g^2 = .002$). Bonferroni post hoc tests revealed significantly lower response times in the joint attention condition compared to both individual ($p < .001$) and control ($p < .001$) conditions, but no significant difference between the individual and the control conditions ($p = .99$).

Mirroring our performance analyses, we then performed two pre-planned pairwise comparisons with permutation-based t-tests on the computed difference in reaction times between one and two beeps for each social condition. We found no significant differences

between individual and joint attention conditions ($t(48) = -1.31$, *corrected* $p = .36$, Cohen's $d = 0.18$), and neither between individual and control conditions, ($t(48) = 0.75$, *corrected* $p = .95$, Cohen's $d = 0.11$).

These results indicate that participants were faster during congruent (2 flashes, 2 beeps) than incongruent (2 flashes, 1 beep) stimuli, and faster in the joint attention condition compared to the individual or control condition, for comparable performance on the flash-counting task.

Signal detection measures

Signal detection theory analysis indicated that sensitivity and criterion bias did not visibly differ across social conditions (Fig. 3, panels C and D, respectively). One-way repeated ANOVAs showed no significant effect of social condition on sensitivity *d'* ($F(2,96) = 0.53$, $p = .58$), and neither on criterion *c* ($F(2,96) = 2.75$, $p = .07$). As in our performance analyses, we performed two pre-planned pairwise comparisons for differences in *d'* and *c*, between individual and joint attention condition, and between individual and control conditions (Table 3). As shown in this table, we found a significant difference in the decision criterion *c* between joint attention ($M = -0.92$, $SD = 0.85$) and individual ($M = -1.09$, $SD = 0.77$) conditions, suggesting that participants were less biased toward the auditory distractor during joint attention, as compared to performing the task individually. This reduced bias was not observed between the individual and control conditions.

Table 3

Pairwise Comparisons of Signal Detection Measures Across Social Conditions for the Fusion Illusion

Measure	Comparison	*t*	*df*	95% CI	Cohen's *d*	Corrected *p*
Sensitivity *d'*	Individual vs joint attention	1.02	48	[-0.11, 0.35]	0.15	0.63
	Individual vs control	0.34	48	[-0.18, 0.25]	0.05	1.00
Criterion *c*	Individual vs joint attention	-2.34	48	[-0.31, -0.02]	0.33	0.04
	Individual vs control	-0.58	48	[-0.19, 0.1]	0.08	1.00

Note. CI = confidence interval; Bonferroni corrected *p* values.

4. Discussion

In this study, we investigated whether the hypothesis that joint attention can boost relative processing of sensory stimuli compared to solo attention (Becchio et al., 2008; Mundy, 2016, 2018; Shteynberg, 2015, 2018) extends to temporal multisensory processing. Specifically, we tested whether engaging in joint attention could reduce temporal audiovisual illusions by enhancing the processing of the jointly attended modality and/or reducing the distraction to the non-attended modality.

While previous work examined the impact of joint attention on stimuli in the tactile or auditory modality with gaze cues displayed on a computer screen (De Jong & Dijkerman, 2019; Nuku & Bekkering, 2010; Soto-Faraco et al., 2005), we investigated the impact of joint

attention on *audiovisual* stimuli and manipulated joint attention by having two participants concurrently know that they are attending to the same target. Using the sound-induced flash illusions, participants counted visual flashes in three social conditions: alone, in pairs sitting in proximity, and with another participant sitting in proximity but with their attention engaged in a different task. In all social conditions, participants could not ignore the jointly presented sounds and were susceptible to both seeing more (i.e., the fission illusion) or fewer flashes (i.e., the fusion illusion) than actually presented. With these findings, we replicate previous studies that focused on individual performance (Andersen et al., 2004; Keil, 2020; Shams et al., 2002). Following the hypothesis that joint attention enhances information encoding and processing of the co-attended stimuli relative to distractors, we predicted that when participants jointly attend and respond to the same visual target stimuli, the sound-induced flash illusions will be reduced. However, people did not perform better nor worse across the different social conditions, in both fission and fusion illusions. These findings suggest that the temporal integration, as measured by the number of flashes reported when presented with incongruent beeps, is robust across all social conditions tested.

Regarding reaction times, people performed faster on the joint attention condition compared to the other social conditions across all stimuli combinations. We suggest that the effect on response times may be due to a social impact on motivation or arousal — a social facilitation effect (Belletier, Normand, & Huguet, 2019). These faster responses, however, did also not result in more incorrect responses. That is, as reported above, the accuracy of reported flashes did not differ.

Using signal detection measures, we found that people's criterion bias was less affected by the auditory beeps for the fusion illusion (i.e., their bias decreased) when engaged in joint attention as contrasted with the individual condition. Such an effect was not observed when comparing the individual and the co-presence control conditions. Interestingly, we only

observed a bias reduction on the fusion illusion while this was not the case for the fission illusion, suggesting that a joint attention manipulation only affects the bias for the fusion illusion but not for the fission illusion. In line with earlier work (Mishra et al., 2008; Watkins at el., 2007; see Hirst et al., 2020, for a review), these findings suggest that the fusion and fission illusions may be mediated by different mechanisms and are thus susceptible to different experimental manipulations.

Recent studies (Tremblay & Nguyen, 2010; Welsh, Reid, Manson, Constable, & Tremblay, 2020) that examined how performing or observing someone's actions affects the fusion illusion may help explain our present fusion illusion effects. In particular, Tremblay & Nguyen (2010) found that the fusion illusion is reduced when participants start a goal-directed reaching movement 50 to 100 ms before the audiovisual stimuli are shown. One likely explanation is that during the earlier stages of a goal-directed movement there is a shift in the relative weighting of sensory information towards vision (Kennedy, Bhattacharjee, Hansen, Reid, & Tremblay, 2015; Manson et al., 2018). In addition, Welsh et al. (2020) report that the fusion illusion is similarly attenuated when participants observe someone else perform the movement, suggesting that participants simulate the performance of the observed action, and thus experience a similar impact on multisensory processing during both action observation and execution. While in our study participants did not engage in any visible motor actions while performing the flash-counting task, one possible interpretation for the reduced bias during the fusion illusion is that the presence of a co-actor engaging in the same task and directing their attention to the same visual target could already at least minimally engage the same mechanisms behind the reduction of the fusion illusion during action execution.

One further proviso is needed to interpret this shift in bias. Witt and colleagues (Witt et al. 2015, 2016) show that a change in *c* does not necessarily reflect a change in non-

perceptual response bias or decision bias, and that the strength of the sound-induced flash illusions should be reflected primarily in the criterion measure. Theoretically, the number of beeps bias perception to detect the same number of flashes (Witt et al. 2015). Knotts & Shams (2016) suggest that both *d'* and *c* may indicate perceptual processes associated with the illusions. Although we cannot straightforwardly determine whether the bias is either purely perceptual or response-based (Witt et al., 2015, 2016), our results indicate that attending to the flashes together with another participant reduces the bias introduced by the sound distractors in the fusion illusion, though this reduction was not enough to effect a significant change in the mean number of flashes reported.

While the present study investigated the impact of joint attention on the sound-induced flash illusion, an earlier study found that a division of labour manipulation, where the participant reported on the number of flashes while a confederate simultaneously reported on the number of beeps, induced a stronger fission illusion compared to performing the task alone (Wahn et al. 2020). The authors suggest that in their social manipulation, the participant's visual attention was divided between the visual flash-counting task and attending to the co-actor which in turn increased the influence of the auditory stimuli and thus the number of perceived fission illusions.

Since participants in a pair performed different tasks in this earlier study, participants likely showed a tendency to represent and monitor the other's performance. For our joint attention manipulation, in contrast, it may not be necessary to co-represent the other person's task, nor monitor their performance, since the other person attended to the same target and had the same task. Given these differences, the participants' visual attention was likely not divided in the present study. This interpretation is in line with evidence showing that performing a task together reduces interference in unisensory Stroop-like tasks only when labour is divided, but not when it is shared (Sellaro, Treccani, & Cubelli, 2018). In their

study, participants had to identify pictures while ignoring distractor words shown concurrently, which induce a semantic interference effect. The disappearance of the interference was observed in the joint task in which participants believed that the co-actor was reading the distractor words (different target), but not in the joint task in which the co-actor was thought to name the colour of the pictures (same target) (Sellaro et al., 2018). Taken together with these studies, the results of the present study indicate that when the participant knows that another actor is taking care of potentially distracting stimuli, a division of labour can be established which affects the participant's performance. But this effect disappears when both participants are attending and responding to the same target stimulus. In short, multisensory integration of temporal stimuli is affected by a division of labour manipulation but not by a joint attention manipulation.

In the present study, we operationalize joint attention as the situation in which two individuals focus their perceptual attention on the same modal target, and both know together that they are so sharing their attention (Siposova & Carpenter, 2019; Tomasello, 1995). This minimal manipulation is sufficient to induce interferences in the case of joint action (Schmitz, Vesper, Sebanz & Knoblich, 2017). Outside the laboratory, however, joint attention comes in varying degrees, depending on how much co-attenders share between them (Siposova & Carpenter, 2019). Future studies could explore whether factors that elicit a stronger feeling of jointness affect multisensory processing. For instance, the feeling of jointness could be enhanced by reciprocal communicative interaction between co-attenders, sharing emotions (e.g., smiling), sharing object-directed action (e.g., joint intentional goals), familiarity or previous relationship between the individuals (e.g. family members, friends, partners). The sense of jointness between participants could also depend on the pay-off structure of the task and the required coordination between them. For example, in the absence of a shared goal, an individual can assign little value in co-representing the other's

performance, even though they are engaging in joint attention. In a case where both co-attenders share the same goal, so that they receive greater rewards when their individual performances are aligned, an individual may thus benefit from co-representing the other's performance and, in turn, their own perceptual processing of the jointly attended target could be thus greatly affected. Future studies could test this proposal, and address the role of different pay-off structures on an individual's multisensory processing during joint attentional tasks.

Finally, our results shore up the limitations of the view that joint attention enhances stimulus information encoding and processing (Becchio et al., 2008; Mundy, 2018; Shteynberg, 2015). While this view explains the effect of joint attention in facilitating mental spatial rotation performance (Böckler et al., 2011), working memory (Gregory & Jackson, 2017; Kim & Mundy, 2012), and enhancing spatial crossmodal attention (De Jong & Dijkerman, 2019; Nuku & Bekkering, 2010), it cannot be straightforwardly applied to the integration of temporal multisensory events. This study provides grounds for future work in comparing the effects of joint attention across temporal and spatial multisensory processes, and map the limitations of the view that joint attention results in greater processing resources to those features of the environment that are co-attended simultaneously.

Acknowledgements

We acknowledge the support of a DFG research fellowship (WA 4153/2-1) awarded to BW. OD was supported by a grant from the Excellence Initiative in the LMU, and a grant from the NOMIS foundation (acronym DISE).

Open Practices Statement

The data generated and analysed during the study are publicly available (https://doi.org/10.17605/OSF.IO/GCUK9). The experiment and analyses were preregistered at the Open Science Framework (https://osf.io/v5gjp).

Author information

LB, BW and OD conceived of the research idea. LB, IG and OD designed the study. LB carried out the experiments, performed the analyses, and drafted the manuscript with consultation from all authors. All authors discussed the results and commented on the manuscript.

References

Aczel, B., Palfi, B., & Szaszi, B. (2017). Estimating the evidential value of significant results in psychological science. *PLOS ONE*, *12*(8), e0182651. Retrieved from https://doi.org/10.1371/journal.pone.0182651

Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, *21*(3), 301–308. https://doi.org/https://doi.org/10.1016/j.cogbrainres.2004.06.004

Battich, L., Fairhurst, M., & Deroy, O. (2020). Coordinating attention requires coordinated senses. *Psychonomic Bulletin & Review, 27*(6), 1126–1138. https://doi.org/10.3758/s13423-020-01766-z

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience*, *7*(11), 1190–1192. https://doi.org/10.1038/nn1333

Becchio, C., Bertone, C., & Castiello, U. (2008). How the gaze of others influences object processing. *Trends in Cognitive Sciences*, *12*(7), 254–258. https://doi.org/10.1016/j.tics.2008.04.005

Belletier, C., Normand, A., & Huguet, P. (2019). Social-facilitation-and-impairment effects: From motivation to cognition and the social brain. *Current Directions in Psychological Science*, *28*(3), 260–265. https://doi.org/10.1177/0963721419829699

Bottema-Beutel, K. (2016). Associations between joint attention and language in autism spectrum disorder and typical development: A systematic review and meta-regression analysis. *Autism Research*, *9*(10), 1021–1035. https://doi.org/10.1002/aur.1624

Böckler, A., Knoblich, G., & Sebanz, N. (2011). Giving a helping hand: effects of joint attention on mental rotation of body parts. *Experimental Brain Research*, *211*(3-4), 531–545. https://doi.org/10.1007/s00221-011-2625-z

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63*(4), i–174. https://doi.org/10.2307/1166214

Cecere, R., Rees, G., & Romei, V. (2015). Individual differences in alpha frequency drive crossmodal illusory perception. *Current Biology*, *25*(2), 231–235. https://doi.org/10.1016/j.cub.2014.11.034

Chelazzi, L., Marini, F., Pascucci, D., & Turatto, M. (2019). Getting rid of visual distractors: the why, when, how, and where. *Current Opinion in Psychology*, *29*, 135–147. https://doi.org/https://doi.org/10.1016/j.copsyc.2019.02.004

Choi, I., Lee, J.-Y., & Lee, S.-H. (2018). Bottom-up and top-down modulation of multisensory integration. *Current Opinion in Neurobiology*, *52*, 115–122. https://doi.org/10.1016/j.conb.2018.05.002

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

De Jong, M. C., & Dijkerman, H. C. (2019). The influence of joint attention and partner trustworthiness on cross-modal sensory cueing. *Cortex*, *119*, 1–11. https://doi.org/10.1016/j.cortex.2019.04.005

DeLoss, D. J., & Andersen, G. J. (2015). Aging, spatial disparity, and the sound-induced flash illusion. *PLOS ONE*, *10*(11), e0143773. https://doi.org/10.1371/journal.pone.0143773

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169. https://doi.org/10.1016/j.tics.2004.02.002

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*(4), 694–724. https://doi.org/10.1037/0033-2909.133.4.694

Gregory, S. E. A., & Jackson, M. C. (2017). Joint attention enhances visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(2), 237–249. https://doi.org/10.1037/xlm0000294

Heed, T., Habets, B., Sebanz, N., & Knoblich, G. (2010). Others' actions reduce crossmodal integration in peripersonal space. *Current Biology*, *20*(15), 1345–1349. https://doi.org/10.1016/j.cub.2010.05.068

Hirst, R. J., McGovern, D. P., Setti, A., Shams, L., & Newell, F. N. (2020). What you see is what you hear: Twenty years of research using the Sound-Induced Flash Illusion. *Neuroscience & Biobehavioral Reviews*, *118*, 759–774. https://doi.org/10.1016/j.neubiorev.2020.09.006

Keil, J. (2020). Double flash illusions: Current findings and future directions. *Frontiers in Neuroscience*, *14*, 298. https://doi.org/10.3389/fnins.2020.00298

Kennedy, A., Bhattacharjee, A., Hansen, S., Reid, C., & Tremblay, L. (2015). Online vision as a function of real-time limb velocity: Another case for optimal windows. *Journal of Motor Behavior*, *47*(6), 465–475. https://doi.org/10.1080/00222895.2015.1012579

Kherad-Pajouh, S. and Renaud, O. (2015). A general permutation approach for analyzing repeated measures ANOVA and mixed-model designs, *Statistical Papers, 56*(4), 947–967. https://doi.org/10.1007/s00362-014-0617-3

Kim, K., & Mundy, P. (2012). Joint attention, social-cognition, and recognition memory in adults. *Frontiers in Human Neuroscience*, *6*, 172. https://doi.org/10.3389/fnhum.2012.00172

Knotts, J. D., & Shams, L. (2016). Clarifying signal detection theoretic interpretations of the Müller–Lyer and sound-induced flash illusions. *Journal of Vision*, *16*(11), 18. https://doi.org/10.1167/16.11.18

Macaluso, E., Noppeney, U., Talsma, D., Vercillo, T., Hartcher-O'Brien, J., & Adam, R. (2016). The curious incident of attention in multisensory integration: Bottom-up vs. Top-down. *Multisensory Research*, *29*(6-7), 557–583. https://doi.org/10.1163/22134808-00002528

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide, 2nd ed.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Manson, G. A., Manzone, D., Grosbois, J. de, Goodman, R., Wong, J., Reid, C., … Tremblay, L. (2018). Let us not play it by ear: Auditory gating and audiovisual perception during rapid goal-directed action. *IEEE Transactions on Cognitive and Developmental Systems*, *10*(3), 659–667. https://doi.org/10.1109/TCDS.2017.2773423

McGovern, D. P., Roudaia, E., Stapleton, J., McGinnity, T. M., & Newell, F. N. (2014). The sound-induced flash illusion reveals dissociable age-related effects in multisensory integration. *Frontiers in Aging Neuroscience*, *6*, 250. https://doi.org/10.3389/fnagi.2014.00250

Michail, G., & Keil, J. (2018). High cognitive load enhances the susceptibility to non-speech audiovisual illusions. *Scientific Reports*, *8*(1), 11530. https://doi.org/10.1038/s41598-018-30007-6

Mishra, J., Martinez, A., & Hillyard, S. A. (2008). Cortical processes underlying sound-induced flash fusion. *Brain Research*, *1242*, 102–115. https://doi.org/10.1016/j.brainres.2008.05.023

Mishra, J., Martínez, A., & Hillyard, S. A. (2010). Effect of Attention on Early Cortical Processes Associated with the Sound-induced Extra Flash Illusion. *Journal of Cognitive Neuroscience*, *22*(8), 1714–1729. https://doi.org/10.1162/jocn.2009.21295

Moorselaar, D. van, & Slagter, H. A. (2020). Inhibition in selective attention. *Annals of the New York Academy of Sciences*, *1464*(1), 204–221. https://doi.org/10.1111/nyas.14304

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Mundy, P. (2016). *Autism and joint attention: Development, neuroscience, and clinical fundamentals*. Guilford Publications.

Mundy, P. (2018). A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *European Journal of Neuroscience*, *47*(6), 497–514. https://doi.org/10.1111/ejn.13720

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science*, *16*(5), 269–274. https://doi.org/10.1111/j.1467-8721.2007.00518.x

Noonan, M. P., Adamian, N., Pike, A., Printzlau, F., Crittenden, B. M., & Stokes, M. G. (2016). Distinct mechanisms for distractor suppression and target facilitation. *The*

Journal of Neuroscience, *36*(6), 1797–1807. https://doi.org/10.1523/JNEUROSCI.2133-15.2016

Nuku, P., & Bekkering, H. (2010). When one sees what the other hears: Crossmodal attentional modulation for gazed and non-gazed upon auditory targets. *Consciousness and Cognition*, *19*(1), 135–143. https://doi.org/10.1016/j.concog.2009.07.012

Odegaard, B., Wozny, D. R., & Shams, L. (2016). The effects of selective and divided attention on sensory precision and integration. *Neuroscience Letters,* 614, 24–28. https://doi.org/10.1016/j.neulet.2015.12.039

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Peirce, J. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (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*(5), 1255–1266. https://doi.org/10.1037/a0018729

Schmitz, L., Vesper, C., Sebanz, N., & Knoblich, G. (2017). Co-representation of others' task constraints in joint action. *Journal of Experimental Psychology: Human Perception and Performance*, 43(8), 1480–1493. https://doi.org/10.1037/xhp0000403

Sellaro, R., Treccani, B., & Cubelli, R. (2018). When task sharing reduces interference: Evidence for division-of-labour in Stroop-like tasks. *Psychological Research*, *84*(2), 327–342. https://doi.org/10.1007/s00426-018-1044-1

Seow, T., & Fleming, S. M. (2019). Perceptual sensitivity is modulated by what others can see. *Attention, Perception, & Psychophysics*, *81*(6), 1979–1990. https://doi.org/10.3758/s13414-019-01724-5

Shams, L., Iwaki, S., Chawla, A., & Bhattacharya, J. (2005). Early modulation of visual cortex by sound: an MEG study. *Neuroscience Letters*, *378*(2), 76–81. https://doi.org/10.1016/j.neulet.2004.12.035

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, *408*(6814), 788–788. https://doi.org/10.1038/35048669

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, *14*(1), 147–152. https://doi.org/https://doi.org/10.1016/S0926-6410(02)00069-1

Shams, L., & Kim, R. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews*, *7*(3), 269–284. https://doi.org/https://doi.org/10.1016/j.plrev.2010.04.006

Shteynberg, G. (2015). Shared attention. *Perspectives on Psychological Science*, *10*(5), 579–590. https://doi.org/10.1177/1745691615589104

Shteynberg, G. (2018). A collective perspective: Shared attention and the mind. *Current Opinion in Psychology*, *23*, 93–97. https://doi.org/10.1016/j.copsyc.2017.12.007

Siposova, B., & Carpenter, M. (2019). A new look at joint attention and common knowledge. *Cognition*, *189*, 260–274. https://doi.org/10.1016/j.cognition.2019.03.019

Soto-Faraco, S., Sinnett, S., Alsius, A., & Kingstone, A. (2005). Spatial orienting of tactile attention induced by social cues. *Psychonomic Bulletin & Review*, *12*(6), 1024–1031. https://doi.org/10.3758/bf03206438

Takeshima, Y. (2020). Emotional information affects fission illusion induced by audio-visual interactions. *Scientific Reports*, *10*(1), 998. https://doi.org/10.1038/s41598-020-57719-y

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410. https://doi.org/10.1016/J.TICS.2010.06.008

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Lawrence Erlbaum.

Tremblay, L., & Nguyen, T. (2010). Real-time decreased sensitivity to an audio-visual illusion during goal-directed reaching. *PLOS ONE*, *5*(1), e8952. https://doi.org/10.1371/journal.pone.0008952

Veale, J. F. (2014). Edinburgh Handedness Inventory – Short Form: A revised version based on confirmatory factor analysis. *Laterality*, *19*(2), 164–177. https://doi.org/10.1080/1357650X.2013.783045

Wahn, B., Keshava, A., Sinnett, S., Kingstone, A., & König, P. (2017). Audiovisual integration is affected by performing a task jointly. *Proceedings of the 39th annual conference of the cognitive science society*, 1296–1301.

Wahn, B., Rohe, T., Gearhart, A., Kingstone, A., & Sinnett, S. (2020). Performing a task jointly enhances the sound-induced flash illusion. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/1747021820942687

Wang, A., Sang, H., He, J., Sava-Segal, C., Tang, X., & Zhang, M. (2019). Effects of cognitive expectation on sound-induced flash illusion. *Perception*, *48*(12), 1214–1234. https://doi.org/10.1177/0301006619885796

Watkins, S., Shams, L., Josephs, O., & Rees, G. (2007). Activity in human V1 follows multisensory perception. *NeuroImage*, *37*(2), 572–578. https://doi.org/10.1016/j.neuroimage.2007.05.027

Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage*, *31*(3), 1247–1256. https://doi.org/10.1016/j.neuroimage.2006.01.016

Welsh, T. N., Reid, C., Manson, G., Constable, M. D., & Tremblay, L. (2020). Susceptibility to the fusion illusion is modulated during both action execution and action observation. *Acta Psychologica*, *204*, 103028. https://doi.org/10.1016/j.actpsy.2020.103028

Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2015). Signal detection measures cannot distinguish perceptual biases from response biases. *Perception*, *44*(3), 289–300. https://doi.org/10.1068/p7908

Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2016). Further clarifying signal detection theoretic interpretations of the Müller–Lyer and sound-induced flash illusions. *Journal of Vision*, *16*(11), 19. https://doi.org/10.1167/16.11.19

Zhang, N., & Chen, W. (2006). A dynamic fMRI study of illusory double-flash effect on human visual cortex. *Experimental Brain Research*, *172*(1), 57–66. https://doi.org/10.1007/s00221-005-0304-7

Chapter 6

General Discussion

> "Joint attention might indeed be where the life of the mind and the social life first meet; and somewhere close to that meeting place is where to find what makes us humans unique among the animals. This should be enough to keep thinking about joint attention." Sebastian Watzl (2012)

In this thesis, I aimed to elucidate the role of perceptual experience in characterising joint attention, to examine how different senses shape joint attention and, conversely, how joint attention can affect perception across modalities. First, I critically assessed the view that joint attention is based on some primitive intersubjective conscious relation, which cannot be explained in terms of individual mental states of each individual. I advanced several arguments against this view and concluded that the theory is not conceptually sound (Paper I). In the second article, I proposed that current philosophical accounts of joint attention often confound normative and psychological explanatory aims. I then proposed an empirically informed account of the openness in joint attention, suggesting that the mutual awareness the co-attenders enjoy is not something they must arrive at, but that it is often implicitly assumed (Paper II). The third article proposed that joint attention is fundamentally a multisensory phenomenon. It showed in detail how non-visual senses make essential contributions to joint attention (Paper III). Building on this proposal, the fourth paper investigated whether joint attention can affect multisensory perception. This study showed that joint attention does not affect an individual's integration of temporal audiovisual stimuli (Paper IV).

In the following sections, I will situate the thesis's research contributions in relation to other areas of philosophy and social cognitive neuroscience, and discuss potential philosophical and scientific implications for multisensory

research and the study of common knowledge in cognitive science. Finally, I will highlight the limitations of the work presented in this thesis.

1 Relationalist theories of joint attention

The relational approach to joint attention is a prominent account of the phenomenon. Its putative advantages over alternative views have made it an attractive approach among philosophers and scientists working on joint attention, social cognition, and communication. As part of this thesis, I have shown that the relationalist's advantages do not hold to scrutiny (Paper I). First, it requires a conceptual understanding of the notion of co-attendance that can lead to a recursion of mental states — the very problem that traditional reductive accounts face. Second, to sustain the unique intersubjective character of perceptual experience that relationalists posit, the theory requires a causal notion of attention monitoring which is either too strict to be realistic or so loose that it cannot be a purely causal notion in the first place. Third, the relational view cannot properly account for erroneous cases of joint attention. Since it is based on an unorthodox disassociation between someone's perceptual experience and their own information about that perceptual experience, it implies that one could be in joint attention without being aware of it — a conclusion that goes against the very feature of joint attention that the view attempts to explain: co-attenders are mutually aware of their joint attention. These arguments raise significant worries for an account for joint attention based on a relational view of perceptual experience and, more generally, for any account that conceives of the mutual awareness in joint attention as a primitive experiential phenomenon.

The relational view of joint attention has become an attractive theoretical position for several cognitive and developmental psychologists, including Moll and Meltzoff's (2011) work on the understanding of others' perspectives, and Hobson and Hobson's (2011) work on the affective aspects of joint attention. Additionally, the relational view has been invoked as a complementary to, and to some extent supportive of, an interactionist approach to social cognition based on embodied, embedded and extended interactive processes (Gallagher, 2010; Gallagher, 2011; León et al., 2019; see also De Jaegher et al., 2010). Similar non-reductive accounts have been proposed for common knowledge (Wilby, 2010) and collective intentionality (Gold & Harbour, 2012; cf. Bacharach, 2006). The conclusion that the relational approach cannot provide a satisfactory theoretical account of joint attention may thus encourage researchers across cognitive and developmental psychology as well

as different areas of philosophy of mind, to consider alternative approaches to joint attention and related intersubjective phenomena.

2 From joint attention to common knowledge

Joint attention is often defined as a mutually "open" relation between co-attenders: they are mutually aware of being so engaged. But how should this openness be characterised? In this thesis, I argued that tensions in current approaches to this question arise from the attempt to address simultaneously two distinct explanatory aims: the normative aim of explaining the role of joint attention in justifying joint endeavours and shared knowledge of the world, and the cognitive aim of explicating the psychological capacities and wherewithal involved in joint attention (Paper II). Current theoretical accounts of joint attention are primarily designed to tackle the normative concerns, and their problems arise when they conflate these concerns with psychological ones. Drawing from evidence in developmental and cognitive psychology, I outlined the case for a cognitive account of joint attention based on a weaker notion of openness and mutual awareness. The arguments advanced in this thesis have a direct implication to debates about common knowledge and the related notion common ground in cognitive science and the philosophy of communication.

Michael Tomasello has previously remarked that child language acquisition is not a logical problem, but an empirical one. He urged that a theory of human linguistic competence should be based less be on analogies to formal languages, and more on empirical research in the cognitive sciences (Tomasello, 2003, 328). Tomasello's plea to distinguish logical matters from empirical ones seems both easy to understand and, alas, easy to miss. The clash between normative logical theories of human behaviours and theories of the biological, psychological and social basis of behaviour in real, actual humans is a recurrent theme in the history of cognitive science and its sibling fields of cognitive neuroscience and neuroeconomics. One salient example is the long-standing distinction between Econs and Humans in behavioural economics where, traditionally, mainstream economists study the former and cognitive scientists the latter. Econs are the agents of classical economics, fully rational and consistent in their preferences. Humans are the real thing (Thaler & Sunstein, 2008; Kahneman, 2011).

We see a similar tension in the philosophy and logic of knowledge. Epistemic and doxastic logics traditionally make assumptions about the rational powers of agents that are patently incompatible with the reasoning abilities of

actual humans. For example, epistemic logics usually assume that knowledge is closed under implication, which results in the psychologically implausible feature that if ϕ is known, then *all* its logical consequences are immediately known (Hintikka, 2005). This is not surprising, given that these logics are meant to model rational epistemic states. They account for the actions and epistemic states of Logons — the logical equivalent to the idealised Econs of behavioural economics. They are not, for the most part, intended to model nor explain human behaviours and the processes behind them (Solaki et al., 2019). In fact, the question of whether humans are to be bound by the norms of deductive epistemic logic is still debatable (Harman, 2002).

There is one object of inquiry in cognitive science on which most researchers, including Tomasello, have not been able to abandon the logical grip: the phenomenon of common knowledge. Drawing from the philosophical and logical literature, many researchers have been driven to the notion that common knowledge must involve something like recursive mindreading — the ability to infer someone else's thoughts about one's own thoughts. The conflation of logical and cognitive aims has given rise to the so-called "problem of common knowledge": how can one account common knowledge without being committed to an infinite regression of mental states? (Wilby, 2010, 86). Thus, Zawidzki (2013) pleads that

> we desperately need a psychologically realistic characterization of mutual/common knowledge, since its importance in human coordination is phenomenologically and empirically obvious. However, standard analyses, in terms of infinite iterations of knowledge attributions (Schiffer, 1972), fail to explain how mutual/common knowledge can play such an important role, due to their psychological implausibility. (Zawidzki, 2013, 124-5)

Tomasello similarly recognises that the problem is still outstanding:

> No one is certain how best to characterize this potentially infinite loop of me monitoring the other, who is monitoring my monitoring of her, and so forth (called recursive mindreading by Tomasello, 2008), but it seems to be part of infants' experience—in some nascent form—from before the first birthday. (Tomasello, 2011, 34-35; cf. Tomasello, 2019, 44)

By and large, debates in the psychology of common knowledge and joint attention have been unable to leave aside the assumption inherent in the pioneering work of Lewis (1969) and Schiffer (1972): the assumption of an idealised notion of rationality based on deductive logic. I have argued that this

assumption is best put aside in analyses of joint attention (Paper II). Since the openness or jointness of joint attention is arguably conceptually similar to the openness of common knowledge (Campbell, 2018; cf. Stalnaker, 2002; Grice, 1957), the same may be argued in respect to later. Following from the arguments presented in this thesis, we should then endeavour to be clear of what we talk about when we talk about common knowledge in cognitive science and philosophy of mind. Are we concerned with the normative aspects of common knowledge, or with the neural and psychological processes behind common knowledge? To maintain clarity, therefore, we should differentiate between common knowledge as a normative phenomenon and the representation of common knowledge in a person's psychology, which is a psychological phenomenon. This later is best analysed on the level of the psychological and neural processing of the individual. This approach, it should be noted, does not require that we accept beforehand the logical theory of common knowledge as a recursion of propositional states.

Given its importance in human coordination, a psychologically realistic characterisation of common knowledge is a key desideratum across the fields of philosophy of action, philosophy of communication, and the social sciences (Zawidzki, 2013). There is, however, little agreement on how to relate the normative aspects of common ground with the psychological states of each individual. The approach I propose in this thesis provides a conceptual starting point to address this issue.

3 Multisensory joint attention

In Paper III of this thesis, I propose that joint attention is fundamentally a multisensory phenomenon, and show in detail how the combination of multiple senses not only facilitates visual coordination, but is even necessary to certain uses of joint attention. This paper suggests that a multisensory framework to joint attention is necessary to examine and study cases when (1) non-visual senses facilitate visual joint attention, (2) people need to coordinate their attention on non-visible properties and events, and (3) to differentiate between joint attending to the sensible properties of a multisensory object and the object as a whole.

This approach brings together social cognition and multisensory research, and has several implications across different fields. The implications for pedagogy, clinical diagnostics, and social robotics have been already treated in the publication. Here, I will note two further important consequences of this work for philosophical debates and cognitive neuroscience.

3.1 Building a shared reality

Joint attention has been considered an essential step for the development of the concept of a shared objective reality, where mind-independent objects are attended in common (Davidson, 1999; Brinck, 2005; Tomasello, 2014; Seemann, 2019; Higgins et al., 2021). The ability to coordinate my attention to an object together with another individual, goes hand in hand with the ability to experience the object as a mind-independent thing separate from myself (Eilan, 2005; Campbell, 2011). On the assumption that joint attention helps us build a shared objective world, if we restrict ourselves to social gaze interactions and vision alone this world would be incredibly impoverished. Moreover, the triadic relation characteristic of joint attention assists in the combination of different properties as belonging to a unitary object, and at the same allows us to differentiate these properties from each other. This cannot occur in a visual-only scenario. Considering the unique contributions of different sensory modalities may thus be necessary for a theoretical account of how a shared objective world can be constructed.

A multisensory view of joint attention has related implications for philosophical debates on ostensive reference. Ludwig Wittgenstein (2009) challenged his readers to point first at a piece of paper, then at its shape, now at its colour, and at its number (which certainly sounds odd). He noted that, although the pointer will have "meant" something different each time they pointed, it cannot be clear from the behavioural aspects of each pointing gesture what is being pointed at — whether the shape, the colour, the fact that it is one piece and not two or three, or just the whole piece of paper. Ostensive gestures are said to be, by themselves, highly ambiguous. Philosophers since Wittgenstein have attempted to provide the necessary normative conditions that the contextual circumstances of an ostensive signal must satisfy, so that they can fix its intended reference. This is particularly pressing when the shared knowledge between interacting individuals is limited, as can be the case in pre-linguistic infants and animals. By and large, this philosophical endeavour has centred on visual gestures and language. Moreover, the ambiguity of ostensive gestures has been usually diagnosed on the basis of isolated and idealised unimodal (usually visual) gestures, such as pointing, and stripped of the emotive and bodily complexities in which such gestures are embedded in reality (Engelland, 2014). The starting point, in other words, is an impoverished version of interactions involving ostensive references, including joint attention. Taking into account the role of multisensory cues and the social strategies they support can help to dispel this impoverished view, and provides resources to address the normative question of how the referent object of an ostensive gesture during joint attention can be negotiated.

3.2 Social neuroscience and multisensory research

The neuroscience of joint attention is a relatively new field. Published in 2005, in the first neuroimaging study to directly investigate joint attention, participants were presented with video clip recordings of a face looking in a direction either congruent or incongruent with the participant's attentional focus (Williams et al., 2005). Experimental paradigms using virtual computer characters, photographs, and pre-recorded clips presented on a screen have been a powerful tool for the study of joint attention and its cognitive effects (Frischen et al., 2007). In particular, these paradigms can elucidate how another person's eye gaze or face orientation affects one's own cognitive and perceptual processes. These paradigms, however, have their limitations. Participants are mere *observers*, approaching the artificial agent or video clip from third-person perspective. But in most social scenarios, however, we jointly attend with real humans. In other words, joint attention comes with a representation of how "we" attend to an object or event together, rather than the unilateral representation of someone else's gaze orientation (Carpenter et al., 1998). Even when using computer avatars, it has been shown that participants approach a task differently when they believe that an avatar is controlled in real-time by a human agent, than when they believe it is controlled by a computer program (Caruana et al., 2017). The recent programme of second-person neuroscience provides an approach to investigate the neural basis of social cognition within the context of a real-time social interaction (Schilbach et al., 2013; Redcay & Schilbach, 2019).

Experimental studies in second-person neuroscience arguably began with three functional magnetic resonance imaging (fMRI) studies on joint attention published a decade ago (Schilbach et al., 2010; Redcay et al., 2010; Saito et al., 2010; see Caruana et al., 2017, for a review). In these studies, participants inside an MRI scanner interacted with a partner by using gaze cues. The partner was either another human presented on a screen via a live video feed (Redcay et al., 2010; Saito et al., 2010), or an anthropomorphic computer avatar whom participants believed was controlled by a confederate in a separate room (Schilbach et al., 2010). Together with further neuroimaging studies using similar paradigms (Redcay et al., 2012; Caruana et al., 2015), this research shows that responding to joint attentional bids is associated with activity in brain regions associated with social cognition, including the medial prefrontal cortex and the posterior temporal sulcus. On the other hand, initiating joint attention is associated with activity in the inferior frontal gyrus, a region also associated with the control of planned actions (Caruana et al., 2017). The results from neuroimaging studies with infants and children overlap, in general, with the results from adult studies (Oberwelland

et al., 2016; Mundy, 2018), although some studies show greater contribution from the ventral stratium in adults (Schilbach et al., 2010; Pfeiffer et al., 2014), a brain area functionally associated with the anticipation of rewards and the processing of reward prediction errors (Daniel & Pollmann, 2014). Taken together, these studies suggest that joint attention recruits perceptual processes, action control processes, and the so-called "mentalising network", which is consistently implicated when reasoning about other people's mental states (Schurz et al., 2014; Caruana et al., 2017; Mundy, 2018; Redcay & Schilbach, 2019).

Second-person neuroscience has been supported by the development of hyperscanning paradigms. Hyperscanning refers to the technique of measuring the neural activity of multiple brains simultaneously (Montague et al., 2002; for reviews see Konvalinka & Roepstorff, 2012; Czeszumski et al., 2020; Nam et al., 2020). Hyperscanning provides a suitable method to study the synchronisation of brain activity during joint attention. It can directly address the question of how a brain engaged in joint attention differs from a brain engaging in the same attentional task but without a co-attender. One of the earlier fMRI studies on joint attention by Saito and colleagues used this technique, as pairs of participants were scanned simultaneously while engaging in real-time through live video feedback (Saito et al., 2010). They found that activity in the right inferior frontal gyrus was synchronised between paired subjects, and suggested that this region may be implicated in sharing visual attentional states with each other. These findings were replicated by a subsequent fMRI hyperscanning study by Koike et al. (2016). This study also reported that pairs of participants display increased eye-blink synchronisation when they have previously engaged in a joint attention task. Hyperscanning studies on joint attention have also reported that the frequency of social eye contact covariates with activity in the right temporal parietal junction during joint attention (Dravida et al., 2020), a brain area commonly associated with mentalising functions (Perner et al., 2006). The field of hyperscanning is young, with several methodological and interpretative issues still debated (e.g. Wass et al., 2020; Novembre & Iannetti, 2021). Nevertheless, this methodology has great potential for elucidating the neural underpinnings of joint attention and other social interactions (Nguyen et al., 2020; Misaki et al., 2021).

It should be noted, however, that most of these studies are based on *visual* joint attention. For example, participants are instructed to attend together to a target either by following the other's gaze, by providing self-initiated gaze cues to the other agent to establish a common gaze direction, or by following an external visual cue (Saito et al., 2010; Koike et al., 2016; Lachat et al., 2012; Redcay et al., 2012). Thus, brain systems typically reported to be involved during

joint attention overlap with systems recruited for eye gaze following and face processing (Mundy, 2018). Moreover, there is evidence that brain structures associated with spatial and temporal processing are also involved in social processing (Parkinson & Wheatley, 2015). Overlap in neural processing may help explain the perceptual effects of performing a task jointly, which suggests that similar brain areas may be involved in social and multisensory processing (Wahn et al., 2020). One of the implications of the proposed framework in this thesis for studying multisensory joint attention, is that it may allow disassociating the neural systems recruited during purely visual joint attention from cases when visual cues are combined with other modal cues, and even when non-visual cues are involved at all.

4 Cognitive influences on the sound-induced flash illusions

One of the outstanding questions in the field of multisensory research is how social factors affect multisensory processes (see, e.g., Wahn et al., 2018). The empirical study in this thesis aims to contribute to this research programme (Paper IV). Previous evidence suggests that visual joint attention facilitates the crossmodal localisation of tactile and auditory targets at the location corresponding to the other's gaze direction (De Jong & Dijkerman, 2019; Nuku & Bekkering, 2010). Yet, it remains unclear whether joint attention only affects spatial localisation or whether it can also modulate the temporal processing of multisensory events. In this study, we investigated whether joint attention affects crossmodal temporal processing using the sound-induced flash illusions, where an incongruent number of visual flashes and auditory beeps induces a single flash to be perceived as two (fission illusion), and two flashes as one (fusion illusion) (Shams et al., 2000; Keil, 2020; Hirst et al., 2020). In an individual condition, each participant performed a flash counting task alone, while in the joint attention condition, two participants sitting next to each other were both performing the flash counting task. A control condition was used to discard the possibility of mere social presence effects, with two participants sitting together but one performing a different task than a flash counting task.

Following the hypothesis that joint attention enhances the stimulus processing of the co-attender target relative to distractors, we predicted that the influence of the auditory distractor over visual perception would diminish, so that the strength of both illusions will be reduced when performing the

task jointly compared to performing it individually. In contrast to this prediction, we found that joint attention did not significantly affect the frequency of the illusions. However, sensitivity and bias analyses revealed that participants' criterion bias in the fusion illusion (two flashes perceived as one) decreased when engaging in joint attention, compared to the individual condition. These results shore up the limitations of the view that joint attention enhances stimulus information encoding and processing (Becchio et al., 2008; Shteynberg, 2015; Mundy, 2018).

The theoretical question addressed in this study is primarily rooted in joint attention research and functional theories of joint attention. However, the paradigm used and the study's results also inform multisensory research and, specifically, add to the understanding of which and how top-down mechanisms can modulate the experience of the sound-induced flash illusions from a first-person perspective.

Several studies have previously tested whether cognitive processes can influence the susceptibility to the illusions by focusing broadly on three types of cognitive influence:

(a) The effect of modality-specific attention: attending to sounds versus attending to flashes (Andersen et al., 2004; Mishra et al., 2010; Odegaard et al., 2016);

(b) the effect of cognitive load, brought about by having the participant concurrently perform a different task (Michail & Keil, 2018); and

(c) the effect of prior expectations about the proportion of illusion-inducing trials (Wang et al., 2019).

It is an open question whether these top-down cognitive modulations of the sound-induced flash illusions are mediated by the same neural mechanisms, although the difference in task demands may suggest that these could be different mechanisms. As reported in this thesis (Paper IV), we found that joint attention does not result in a lower frequency of illusions. Although the rate of the illusions was not affected, our results also indicate that attending to the flashes together with another participant reduces the bias introduced by the sound distractors in the fusion illusion. Given the different task-demands between social and non-social flash-counting tasks, these results support the suggestion that the mechanisms behind the reduced bias are, to some extent, different than the mechanisms behind the three previously studied top-down cognitive modulations of the illusions: (a) modality-specific attention, (b) cognitive load, and (c) expectations about trial proportions. These results encourage further work to identify different mechanisms behind the attentional

modulations of the sound-induced flash illusions, and thus contribute to our understanding of how these illusions are processed in the brain. Together with evidence that a division of labour manipulation induces a higher fission illusion rate (Wahn et al., 2020), and that joint attention facilitates spatial crossmodal detection (De Jong & Dijkerman, 2019; Nuku & Bekkering, 2010), our results provide grounds and motivation for future research assessing how different social factors influence multisensory processes.

5 Limitations

As an object of academic study, the phenomenon of joint attention is situated at the crossroads between research on individual perception and attention, and research on social cognition and interaction. The study of joint attention is thus largely multidisciplinary. Joint attention has been approached by developmental and comparative psychologists, by cognitive neuroscientists and, more recently, by philosophers with various agendas. This plurality of perspectives is certainly welcome. However, it has the consequence that the current debate on joint attention can seem seldom unified.

The research in this thesis partly reflects this plurality, and thus its corollary fragmented character. The thesis is composed of four collected papers that together address the role of perceptual experience and multiple sensory modalities in joint attention. Given that each paper is a separate stand-alone article, however, each has its own, more narrow, agenda. Moreover, I have approached the subject of perceptual joint attention from an interdisciplinary perspective, using both philosophical and scientific methodologies. This fragmented approach has its advantages. It has allowed me to focus on selected aspects of joint attention and make specific contributions to various debates. Thus, Papers I and II are situated within the theoretical debate explaining the jointness of joint attention. In contrast, Papers III and IV largely take the jointness question for granted, and focus instead on the role of multiple sensory modalities and their integration during joint attention.

Unfortunately, this fragmented approach has its limitations, as each paper pursues a different sub-objective, using different methodologies. Therefore, this thesis does not present a fully unified account of joint attention across all chapters. For example, while the merits of relationalism are of importance to psychologists, the empirical study presented in this thesis does not directly depend on whether that particular theoretical view is sound or not. It is hoped, however, that this thesis shows how research in joint attention can be sustained across disciplinary boundaries, and offer contributions to several fields of research.

Another important limitation of this thesis concerns my positive proposal of how to analyse the mutual awareness in joint attention (Paper II). I have proposed that mutual awareness is neither a primitive state nor a reductive intersubjective relation that co-attenders must arrive at, but that it is often implicitly assumed. We have to (un)learn that other people may not attend to the same things we attend, or may not share the same perceptual knowledge we are currently enjoying. While I provide empirical work in developmental and cognitive psychology to support this view, it is nevertheless underdeveloped. It does not yet fully address what I termed the cognitive question: What cognitive capacities and mental processes or understanding are involved in joint attention?

We need a full account of the minimal capacities necessary to assume that one is mutually sharing attention to the same object with others. I suggested that a minimal requirement for joint attention is that one be able to recognise and respond to the set cues that that other person provides by implicitly assuming that they engage with the object in the same way. But what are these set of cues, precisely? Tackling this question is in large part an empirical project, but one that will benefit from conceptual analysis. Similarly, I have suggested that the assumption of mutual awareness often occurs pre-reflectively and need not be conscious. But more needs to be said about this. How can this assumption occur unconsciously, and still impinge on a person's sense of interpersonal engagement? The positive proposal advanced in this thesis is therefore preliminary and unfinished. It is ripe, in other words, to serve as the basis for my postdoctoral research.

Finally, one further important aspect that has been left untouched in this thesis is the relation between the philosophy of joint attention and the philosophy of attention more generally. The empirical study of joint attention is largely based on the neuroscientific view that attention is fundamentally a type of neuronal or computational mechanism or process. For example, according to the dynamic spotlight or zoom lens model, selective attention involves the enhancement of information processing of a limited area or stimulus in the environment, while decreasing processing of other information (Eriksen & James, 1986; Klein & Lawrence, 2012). Following this conception, Mundy suggests that "joint attention is a socially coordinated spotlight that results in enhanced information processing of a common point of references for social partners in ways that solo attention does not" (Mundy, 2018).

Several philosophers, however, have recently proposed theories of the nature of attention that go beyond its role in subpersonal processing (see Watzl, 2011, for a review). Some of these theories conceive of attention as fundamentally a personal level mental activity. According to these views, the

phenomenal character of attention becomes crucial to explain its functional role (Smithies, 2011; Watzl, 2017). While these theories have been primarily focused on an individual's attention, a theoretical treatment of joint attention will certainly be enriched by making an explicit connection to the different philosophies of attention.

6 Concluding remarks

In this doctoral thesis, I used philosophical and empirical methods to examine the role of perceptual experience in joint attention, including in cases involving multiple sensory modalities. The thesis' collected papers address this aim through two related sub-objectives. First, to clarify the role of mutual awareness and perceptual experience in characterising joint attention. Second, to propose a functional framework to assess multisensory contributions to establishing and maintaining joint attention and, in turn, how joint attention may affect multisensory perception.

Research in philosophy of mind and cognitive psychology has traditionally been centred on the individual. We tend to think of the mind in terms of the individual thoughts and skills of a self-contained and self-sufficient person, who gathers and processes information from the world outside her head. Yet, at the same time, we know from everyday experience that we are deeply enmeshed in social communities, which can, and does, influence our thoughts and cognitive skills in fundamental ways. While these two approaches to the mind are often seen as opposed to each other, the need to bridge the gap between individual cognition and social interaction is increasingly recognised by neuroscientists, psychologists, and philosophers (Bahrami et al., 2010; Schilbach et al., 2013; Wahn et al., 2018; Przyrembel et al., 2012; Redcay & Schilbach, 2019). The research collected in this thesis forms part of this larger trend.

Bibliography

Adamson, L. B., Bakeman, R., Suma, K., & Robins, D. L. (2019). An expanded view of joint attention: Skill, engagement, and language in typical development and autism. *Child Development*, 90(1), e1–e18.

Admoni, H. & Scassellati, B. (2017). Social eye gaze in human-robot interaction: A review. *Journal of Human-Robot Interaction*, 6(1), 25–63.

Akhtar, N. & Gernsbacher, M. A. (2008). On privileging the role of gaze in infant social cognition. *Child Development Perspectives*, 2(2), 59–65.

Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, 21(3), 301–308.

Bacharach, M. (2006). *Beyond Individual Choice: Teams and Frames in Game Theory*. Princeton, NJ: Princeton University Press.

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, 329(5995), 1081–1085.

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.

Becchio, C., Bertone, C., & Castiello, U. (2008). How the gaze of others influences object processing. *Trends in Cognitive Sciences*, 12(7), 254–258.

Böckler, A., Knoblich, G., & Sebanz, N. (2011). Giving a helping hand: effects of joint attention on mental rotation of body parts. *Experimental Brain Research*, 211(3-4), 531–545.

Botero, M. (2016). Tactless scientists: Ignoring touch in the study of joint attention. *Philosophical Psychology*, 29(8), 1200–1214.

Brinck, I. (2005). Critical review of John Campbell: Reference and consciousness. *Theoria*, 3, 266–276.

Bruner, J. S. (1974). From communication to language: A psychological perspective. *Cognition*, 3(3), 255–287.

Bruner, J. S. (1995). From joint attention to the meeting of minds. In C. Moore & P. J. Dunham (Eds.), *Joint Attention: Its Origins and Role in Development* (pp. 1–14). Hillsdale, NJ: Lawrence Erlbaum.

Bruner, J. S. (1998). Routes to reference. *Pragmatics & Cognition*, 6(1-2), 209–227.

Burge, T. (2005). Disjunctivism and perceptual psychology. *Philosophical Topics*, 33(1), 1–78.

Burge, T. (2010). *Origins of Objectivity*. Oxford: Oxford University Press.

Calabi, C. (2008). Winks, sighs and smiles? Joint attention, common knowledge and ephemeral groups. In H. B. Schmid, K. Schulte-Ostermann, & N. Psarros (Eds.), *Concepts of Sharedness: Essays on Collective Intentionality* (pp. 41–58). Frankfurt: De Gruyter.

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

Campbell, J. (2005). Joint attention and common knowledge. In N. M. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint Attention: Communication and Other minds. Issues in Philosophy and Psychology* (pp. 287–297). Oxford: 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. 415–30). Cambridge, MA: MIT Press.

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

Carey, S. (2009). *The Origin of Concepts*. Oxford: Oxford University Press.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63(4), i–174.

Caruana, N., Brock, J., & Woolgar, A. (2015). A frontotemporoparietal network common to initiating and responding to joint attention bids. *NeuroImage*, 108, 34–46.

Caruana, N., McArthur, G., Woolgar, A., & Brock, J. (2017). Simulating social interactions for the experimental investigation of joint attention. *Neuroscience & Biobehavioral Reviews*, 74, 115–125.

Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., Schreiber, F., Rendon, Z. Z., & König, P. (2020). Hyperscanning: A valid method to study neural inter-brain underpinnings of social interaction. *Frontiers in Human Neuroscience*, 14, 39.

Daniel, R. & Pollmann, S. (2014). A universal role of the ventral striatum in reward-based learning: Evidence from human studies. *Neurobiology of Learning and Memory*, 114, 90–100.

Davidson, D. (1999). The emergence of thought. *Erkenntnis*, 51(1), 511–521.

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, 14(10), 441–7.

De Jong, M. C. & Dijkerman, H. C. (2019). The influence of joint attention and partner trustworthiness on cross-modal sensory cueing. *Cortex*, 119, 1–11.

Depowski, N., Abaya, H., Oghalai, J., & Bortfeld, H. (2015). Modality use in joint attention between hearing parents and deaf children. *Frontiers in Psychology*, 6, 1556.

Dravida, S., Noah, J. A., Zhang, X., & Hirsch, J. (2020). Joint attention during live person-to-person contact activates rTPJ, including a sub-component associated with spontaneous eye-to-eye contact. *Frontiers in Human Neuroscience*, 14, 201.

Eilan, N. (2005). Joint attention, communication, and mind. In N. Elian, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint Attention: Communication and Other Minds. Issues in Philosophy and Psychology* (pp. 1–33). Oxford: Oxford University Press.

Eilan, N. (2015). Joint Attention and the Second Person (draft). Retrieved from https://warwick.ac.uk/fac/soc/philosophy/people/eilan/jaspup.pdf.

Eilan, N., Hoerl, C., McCormack, T., & Roessler, J., Eds. (2005). *Joint Attention: Communication and Other minds. Issues in Philosophy and Psychology*. Oxford: Oxford University Press.

Engelland, C. (2014). *Ostension*. Cambridge, MA: The MIT Press.

Eriksen, C. W. & James, J. D. S. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, 40(4), 225–240.

Foxx, R. M. (1977). Attention training: The use of overcorrection avoidance to increase the eye contact of autistic and retarded children. *Journal of Applied Behavior Analysis*, 10(3), 1311211.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, 133(4), 694–724.

Gallagher, S. (2010). Joint attention, joint action, and participatory sense making. *Revue de Phénoménologie*, 18, 111–124.

Gallagher, S. (2011). Interactive coordination in joint attention. In A. Seemann (Ed.), *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience* (pp. 293–305). Cambridge, MA: MIT Press.

Gold, N. & Harbour, D. (2012). Cognitive primitives of collective intentions: Linguistic evidence of our mental ontology. *Mind and Language*, 27(2), 109–134.

Gregory, S. E. A. & Jackson, M. C. (2017). Joint attention enhances visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(2), 237–249.

Grice, H. P. (1957). Meaning. *Philosophical Review*, 66(3).

Harman, G. (2002). Internal critique: A logic is not a theory of reasoning and a theory of reasoning is not a logic. In D. M. Gabbay, R. H. Johnson, H. J. Ohlbach, J. B. T. S. i. L. Woods, & P. Reasoning (Eds.), *Handbook of the Logic of Argument and Inference*, volume 1 (pp. 171–186). Amsterdam: Elsevier.

Higgins, E. T., Rossignac-Milon, M., & Echterhoff, G. (2021). Shared reality: From sharing-is-believing to merging minds. *Current Directions in Psychological Science*, 30(2), 103–110.

Hintikka, J. (2005). *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. London: King's College London Publications.

Hirst, R. J., McGovern, D. P., Setti, A., Shams, L., & Newell, F. N. (2020). What you see is what you hear: Twenty years of research using the Sound-Induced Flash Illusion. *Neuroscience & Biobehavioral Reviews*, 118, 759–774.

Hobson, P. & Hobson, J. (2011). Joint attention or joint engagement? Insights from autism. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 115–136). Cambridge, MA: MIT Press.

Kahneman, D. (2011). *Thinking Fast and Slow*. New York, NY: Farrar, Straus and Giroux.

Keil, J. (2020). Double flash illusions: Current findings and future directions. *Frontiers in Neuroscience*, 14, 298.

Keller, H. (1903). *The Story of My Life*. New York, NY: Doubleday.

Kim, K. & Mundy, P. (2012). Joint attention, social-cognition, and recognition memory in adults. *Frontiers in Human Neuroscience*, 6, 172.

Klein, R. M. & Lawrence, M. A. (2012). On the modes and domains of attention. In M. I. Posner (Ed.), *Cognitive Neuroscience of Attention, 2nd ed.* (pp. 11–28). New York, NY: Guilford Press.

Koike, T., Tanabe, H. C., Okazaki, S., Nakagawa, E., Sasaki, A. T., Shimada, K., Sugawara, S. K., Takahashi, H. K., Yoshihara, K., Bosch-Bayard, J., & Sadato, N. (2016). Neural substrates of shared attention as social memory: A hyperscanning functional magnetic resonance imaging study. *NeuroImage*, 125, 401–412.

Konvalinka, I. & Roepstorff, A. (2012). The two-brain approach: How can mutually interacting brains teach us something about social interaction? *Frontiers in Human Neuroscience*, 6, 215.

Lachat, F., Hugueville, L., Lemaréchal, J.-D., Conty, L., & George, N. (2012). Oscillatory brain correlates of live joint attention: A dual-EEG study. *Frontiers in Human Neuroscience*, 6, 156.

León, F., Szanto, T., & Zahavi, D. (2019). Emotional sharing and the extended mind. *Synthese*, 196(12), 4847–4867.

Lewis, D. (1969). *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Martin, M. (2004). The limits of self-awareness. *Philosophical Studies*, 120(1-3), 37–89.

Michail, G. & Keil, J. (2018). High cognitive load enhances the susceptibility to non-speech audiovisual illusions. *Scientific Reports*, 8(1), 11530.

Misaki, M., Kerr, K. L., Ratliff, E. L., Cosgrove, K. T., Simmons, W. K., Morris, A. S., & Bodurka, J. (2021). Beyond synchrony: the capacity of fMRI hyperscanning for the study of human social interaction. *Social Cognitive and Affective Neuroscience*, 16(1-2), 84–92.

Mishra, J., Martínez, A., & Hillyard, S. A. (2010). Effect of attention on early cortical processes associated with the sound-induced extra flash illusion. *Journal of Cognitive Neuroscience*, 22(8), 1714–1729.

Moll, H. & Meltzoff, A. N. (2011). Joint attention as the fundamental basis of understanding perspectives. In A. Seemann (Ed.), *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience* (pp. 393–414). Cambridge, MA.: MIT Press.

Montague, P., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., Wiest, M. C., Karpov, I., King, R. D., Apple, N., & Fisher, R. E. (2002). Hyperscanning: Simultaneous fMRI during linked social interactions. *NeuroImage*, 16(4), 1159–1164.

Moore, C. & Dunham, P. J. (1995). Current themes in research of joint attention. In C. Moore & P. J. Dunham (Eds.), *Joint Attention: Its Origins and Role in Development* (pp. 15–28). Hillsdale, NJ: Lawrence Erlbaum.

Mundy, P. (2016). *Autism and Joint Attention: Development, Neuroscience, and Clinical Fundamentals*. New York, NY: Guilford Publications.

Mundy, P. (2018). A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *European Journal of Neuroscience*, 47(6), 497–514.

Nam, C. S., Choo, S., Huang, J., & Park, J. (2020). Brain-to-brain neural synchrony during social interactions: A systematic review on hyperscanning studies. *Applied Sciences*, 10(19), 6669.

Nguyen, T., Bánki, A., Markova, G., & Hoehl, S. (2020). Studying parent-child interaction with hyperscanning. In S. Hunnius & M. Meyer (Eds.), *New Perspectives on Early Social-cognitive Development*, volume 254 (pp. 1–24). Amsterdam: Elsevier.

Novembre, G. & Iannetti, G. D. (2021). Hyperscanning alone cannot prove causality. Multibrain stimulation can. *Trends in Cognitive Sciences*, 25(2), 96–99.

Nuku, P. & Bekkering, H. (2010). When one sees what the other hears: Cross-modal attentional modulation for gazed and non-gazed upon auditory targets. *Consciousness and Cognition*, 19(1), 135–43.

Núñez, M. (2014). *Joint Attention in Deafblind Children: A Multisensory Path Towards a Shared Sense of the World*. Technical report, Sense, London.

Oberwelland, E., Schilbach, L., Barisic, I., Krall, S., Vogeley, K., Fink, G., Herpertz-Dahlmann, B., Konrad, K., & Schulte-Rüther, M. (2016). Look into my eyes: Investigating joint attention using interactive eye-tracking and fMRI in a developmental sample. *NeuroImage*, 130, 248–260.

Odegaard, B., Wozny, D. R., & Shams, L. (2016). The effects of selective and divided attention on sensory precision and integration. *Neuroscience Letters*, 614, 24–28.

Parkinson, C. & Wheatley, T. (2015). The repurposed social brain. *Trends in Cognitive Sciences*, 19(3), 133–141.

Peacocke, C. (2005). Joint attention: Its nature, reflexivity, and relation to common knowledge. In N. M. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint Attention: Communication and Other Minds. Issues in Philosophy and Psychology* (pp. 298–324). Oxford: Oxford University Press.

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neuroscience*, 1(3-4), 245–258.

Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., & Vogeley, K. (2014). Why we interact: On the functional role of the striatum in the subjective experience of social interaction. *NeuroImage*, 101, 124–137.

Przyrembel, M., Smallwood, J., Pauen, M., & Singer, T. (2012). Illuminating the dark matter of social neuroscience: Considering the problem of social interaction from philosophical, psychological, and neuroscientific perspectives. *Frontiers in Human Neuroscience*, 6, 190.

Rakoczy, H. (2018). Development of collective intentionality. In M. Jankovic & K. Ludwig (Eds.), *The Routledge Handbook of Collective Intentionality* (pp. 407–419). New York, NY: Routledge.

Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D., & Saxe, R. (2010). Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. *NeuroImage*, 50(4), 1639–1647.

Redcay, E., Kleiner, M., & Saxe, R. (2012). Look at this: The neural correlates of initiating and responding to bids for joint attention. *Frontiers in Human Neuroscience*, 6, 169.

Redcay, E. & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, 20(8), 495–505.

Saito, D., Tanabe, H., Izuma, K., Hayashi, M., Morito, Y., Komeda, H., Uchiyama, H., Kosaka, H., Okazawa, H., Fujibayashi, Y., & Sadato, N. (2010). "Stay tuned": Inter-individual neural synchronization during mutual gaze and joint attention. *Frontiers in Integrative Neuroscience*, 4, 127.

Scaife, M. & Bruner, J. (1975). The capacity for joint visual attention in the infant. *Nature*, 253, 265–266.

Schiffer, S. R. (1972). *Meaning*. Oxford: Clarendon Press.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, 36(04), 393–414.

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., Shah, N. J., Fink, G. R., & Vogeley, K. (2010). Minds made for sharing: Initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience*, 22(12), 2702–2715.

Schmitz, M. (2015). Joint attention and understanding others. *Synthesis Philosophica*, 29(2), 235–251.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, 42, 9–34.

Seemann, A. (2011a). Introduction. In A. Seemann (Ed.), *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience* (pp. 1–17). Cambridge, MA: MIT Press.

Seemann, A., Ed. (2011b). *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*. Cambridge, MA: MIT Press.

Seemann, A. (2011c). 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). Cambridge, MA: MIT Press.

Seemann, A. (2019). *The Shared World: Perceptual Common Knowledge, Demonstrative Communication, and Social Space*. Cambridge, MA: MIT Press.

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.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14(1), 147–152.

Shteynberg, G. (2015). Shared attention. *Perspectives on Psychological Science*, 10(5), 579–590.

Smithies, D. (2011). Attention is rational-access consciousness. In C. Mole, D. Smithies, & W. Wu (Eds.), *Attention: Philosophical and Psychological Essays*. Oxford: Oxford University Press.

Solaki, A., Berto, F., & Smets, S. (2019). The logic of fast and slow thinking. *Erkenntnis*.

Stalnaker, R. (2002). Common Ground. *Linguistics and Philosophy*, 25(5-6), 701–721.

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology*, 55(1), 96–109.

Thaler, R. H. & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.

Tomasello, M. (2003). *Constructing a Language*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.

Tomasello, M. (2011). Human culture in evolutionary perspective. In M. J. Gelfand, C. Chiu, & Y. Hong (Eds.), *Advances in Culture and Psychology* (pp. 5–51). Oxford: Oxford University Press.

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

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

Tomasello, M. & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 57(6), 1454–1463.

Travis, C. (2004). The silence of the senses. *Mind*, 113(449), 57–94.

Wahn, B., Kingstone, A., & König, P. (2018). Group benefits in joint perceptual tasks: A review. *Annals of the New York Academy of Sciences*, 1426(1), 166–178.

Wahn, B., Rohe, T., Gearhart, A., Kingstone, A., & Sinnett, S. (2020). Performing a task jointly enhances the sound-induced flash illusion. *Quarterly Journal of Experimental Psychology*.

Wang, A., Sang, H., He, J., Sava-Segal, C., Tang, X., & Zhang, M. (2019). Effects of cognitive expectation on sound-induced flash illusion. *Perception*, 48(12), 1214–1234.

Wass, S. V., Whitehorn, M., Marriott Haresign, I., Phillips, E., & Leong, V. (2020). Interpersonal neural entrainment during early social interaction. *Trends in Cognitive Sciences*, 24(4), 329–342.

Watzl, S. (2011). The nature of attention. *Philosophy Compass*, 6(11), 842–853.

Watzl, S. (2012). Review of "Joint Attention: New Develoments in Psychology, Philosophy of Mind, and Social Neuroscience" edited by Axel Seemann. *Notre Dame Philosophical Reviews*.

Watzl, S. (2017). *Structuring Mind: The Nature of Attention and how it Shapes Consciousness*. Oxford: Oxford University Press.

Wilby, M. (2010). The simplicity of mutual knowledge. *Philosophical Explorations*, 13(2), 83–100.

Williams, J. H., Waiter, G. D., Perra, O., Perrett, D. I., & Whiten, A. (2005). An fMRI study of joint attention experience. *NeuroImage*, 25(1), 133–140.

Wittgenstein, L. (2009). *Philosophical Investigations*. Oxford: Wiley-Blackwell.

Yang, G.-Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., Jacobstein, N., Kumar, V., McNutt, M., Merrifield, R., Nelson, B. J., Scassellati, B., Taddeo, M., Taylor, R., Veloso, M., Wang, Z. L., & Wood, R. (2018). The grand challenges of Science Robotics. *Science Robotics*, 3(14), eaar7650.

Yu, C. & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS ONE*, 8(11), e79659.

Zawidzki, T. W. (2013). *Mindshaping: A New Framework for Understanding Human Social Cognition*. Cambridge, MA: MIT Press.

Eidesstattliche Versicherung / Affidavit

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation "The Nature of Joint Attention: Perception and Other Minds" selbstständig angefertigt habe, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

I hereby confirm that the dissertation "The Nature of Joint Attention: Perception and Other Minds" is the result of my own work and that I have only used sources or materials listed and specified in the dissertation.

München, den 03.05.2021
Munich, date 03.05.2021

.....
Lucas Battich

Author contributions

Battich, L. and Geurts, B. (2020). Joint attention and perceptual experience. *Synthese*, doi: 10.1007/s11229-020-02602-6.

L.B. conceived of the research idea and arguments presented in the paper, wrote the original draft, and revised the paper with help from B.G.

Battich, L. (manuscript). Opening up the openness of joint attention.

L.B. is the first and sole author of this manuscript.

Battich, L., Fairhurst, M., and Deroy, O. (2020). Coordinating attention requires coordinated senses. *Psychonomic Bulletin & Review, 27*(6), 1126–1138. doi: 10.3758/s13423-020-01766-z.

L.B., M.F. and O.D. conceived of the research idea. L.B. wrote the original draft, and revised the paper with help from all other authors.

Battich, L., Garzorz, I., Wahn, B., and Deroy, O. (forthcoming 2021). The impact of joint attention on the sound-induced flash illusions. *Attention, Perception and Psychophysics*.

L.B., B.W. and O.D. conceived of the research idea. L.B., I.G. and O.D. designed the study. L.B. carried out the experiments, analysed the data, and wrote the paper with help from all authors.

.....	.....
Lucas Battich	Prof. Dr. Ophelia Deroy

Munich, 03 May 2021