Tracking objects, Tracking agents

Nicolas J. Bullot, Patrick Rysiew

Tracking objects, Tracking agents

Nicolas J. Bullot

Institut Jean Nicod (CNRS, EHESS, ENS), 1 bis, avenue de Lowendal, 75007 Paris, France
LPPA, Collège de France, 11, place Marcelin Berthelot, 75231 Paris Cedex 05, France

Patrick Rysiew

Department of Philosophy
University of British Columbia
Vancouver BC V6T-1Z1 Canada

1 Introduction

Animals and humans have to keep track of *individuals* in their environment, both in perception (sensorimotor tracking) and in cognition (e.g., spatio-temporal localization and linguistic reference via memory, communication and reasoning). Items that are typical targets for tracking are things such as stationary physical objects (e.g., rocks, plants, trees, buildings, or attached artifacts), moving physical objects (e.g., animals, certain artifacts) and human beings. All such items are located in a spatial environment, persist over time, and are – or at least closely related to, in the case of conspecifics' bodies – *physical objects* that respect non trivial *objecthood criteria* such as having cohesive parts, following continuous spatio-temporal paths, or possessing causal powers and dispositions. Perceptually tracking these objects through space and over time requires possessing sensorimotor systems (e.g., the oculomotor system, the visuo-haptic system or the auditory system) that can anchor into and smoothly pursue objects' properties. Nonetheless, one may suspect that tracking *intentional agents* (i.e. creatures to whom it is natural to attribute intentional states such as beliefs, plans, desires, and who may even participate in shared intentionality), as opposed to *physical objects without mental states* (i.e. objects which are not intentional agents) exploits or requires further abilities and strategies. In particular, at least for humans, tracking conspecifics amounts to tracking intentional agents. This raises the question of how the perceptual tracking of non-intentional objects relates to the keeping track of intentional agents. Here, we propose an extension and augmentation of recent work on object-tracking to the tracking of intentional agents. Based on the examination of the elementary procedures available for pursing agency, our principal suggestions are as follows.

First, identifying intentional agents is significantly dependent upon the perceptual abilities of physical-object tracking, and might therefore be explained by the 'object-file' hypothesis (this hypothesis is explained in section 3), which we suggest to use in the study of the tracking of intentional agents (section 4). In the most elementary case, humans track intentional agents *as physical objects*: they track such agents by tracking their bodies (section 5). Even though this kind of tracking is insufficient for keeping track of human individuals '*as*' intentional agents (and explaining their behavior with an intentional stance), it may suffice to explain a number of interacting and situated behaviors in social contexts *with* intentional agents. Second, however, tracking intentional agents '*as*' *intentional agents* requires additional capacities for detecting and understanding intentional states and certain further properties which creatures with such states can exhibit (section 6) – e.g., (ir/)rationality, and the capacity to participate in shared intentionality. We note, however, that tracking a human individual as an intentional agent may require an appeal to specific perceptual cues and may even recruit basic sensorimotor skills – such as the detection of biological motions – whose tracking might be independent of the understanding of conspecifics' mental states. For reasons of parsimony and computational economy, unless we have reason to think that there is a separate system devoted to tracking intentional agents, we should suppose that the same mechanism used to track physical objects is recruited for tracking intentional agents.

2 Keeping track of token individuals and the uniqueness problem for object tracking

We shall assume a *realist ontology* of the material world. According to this ontology, the world is composed by

individual physical objects/bodies – that is, tokens of external mind-independent individuals with identity conditions (e.g., Strawson, 1959). If this ontology is correct, the identity conditions of physical objects (including the bodies of human agents or animal organisms) can be explained on the basis of mind-independent causal properties and composition principles – such as chemical composition and genetic properties for living organisms, continuity criteria for objects in general, and so on. Let us call *objecthood properties* those properties – such as cohesion, continuity and impenetrability – which are criteria of physical objecthood, and which are directly relevant to the identification and re-identification of particular (physical) objects. Objecthood properties and criteria have been studied and discussed in both philosophy and psychology.

We will accept also an *object-dependant epistemology*. According to this epistemology, human knowledge is based on skills and singular representations/contents that enable us to single out and track mind-independent objects.[1] In order to interact with objects, to access their properties, to ascribe to them abstract properties or to reason about them, humans need to be able to pick out individual entities, fix a referential link with them and maintain that link over time. However, because the available properties or descriptions of an individual *x* (at a given moment) are not always sufficient to single *x* out and refer to *x* as the same individual, cognizers are confronted with *the uniqueness problem of object tracking* in many situations: For any given unique physical object *x*, how can the cognizer keep track over space and time of *x*'s 'sameness and uniqueness'? The object-dependant epistemology has thus to postulate that cognizers possess capacities which enable them to track perceptually *x* and to keep track cognitively over time of *x* as the same continuing individual – i.e. as a function of *x*'s sameness and uniqueness. For these capacities to be the basis of keeping track of *x*'s uniqueness, they must enable the cognizer to *distinguish* over time *x* from all the other individuals of the same kind. When used for conceptual thinking, this ability is what philosophers have referred to as singular representations and jugdments fulfilling the 'Russell's principle' (Evans, 1982) or the 'know which' requirement (e.g., McDowell, 1990; McDowell, 1998; Peacocke, 1991; Strawson, 1997). According to Pylyshyn (2003), there is also reason to think that such an ability must be ascribed to the visual system itself.[2]

---

[1] A pionering philosophical analysis of human knowledge based on the cognition of individual objects can be found in Strawson (1959). Philosophers such as Evans (1981; 1982), Perry (2001b), Campbell (1994; 2002), Clark (2000) have developed this view of cognition in the philosophy of mind and perception. Views expressed by philosophers of language such as Kripke (1980), McDowell (1984; 1990), or Recanati (1993) may also be consistent with it. Lastly, recent work in the psychology of perception by Kahneman et al. (1984; 1992), Spelke et al. (1990; 1995) and Pylyshyn (2003) may provide empirical ground for an understanding of the cognition of physical objects.

[2] Compare the following assertions. According to Evans on Russell's Principle, "The principle is that a subject cannot make a judgment about something unless he knows which object his judgment is about. (…) [*Evans*] shall suppose that the knowledge which it requires is what might be called *discriminative knowledge*: the subject must have a capacity to distinguish the object of his judgment from all other things. (…) we have the idea of certain sufficient conditions for being able to discriminate an object from all other things: for example, when one is perceiving it at the present time; when one is recognize it if presented with it; and when one knows distinguishing facts about it" (Evans, 1982: 89). According to Strawson (1997: 22-23), the philosophical slogan 'no entity without identity' can be interpreted as an "admirable maxim" which states, "There is nothing you can sensibly talk about without knowing, at least in principle, how it might be identified." According to McDowell (1990: 256-7) "The essential background for Evans's account of demonstrative mode of presentation is the principle that to entertain a thought, one must know what it would be for the thought to be true. It is an application of this principle that if there is some particular object which must be a certain way for the thought to be true, the subject must know which object it is. This requirement – the 'know which' requirement – is what Evans calls 'Russell's Principle'." Finally, according to Pylyshyn, "(…) if the visual system is to do something concerning some visual object, it must in some sense know *which* object it is doing it *to*" (Pylyshyn, 2003: 201).

3 The notion of 'object file' as an explanatory notion in understanding the cognitive relation to the object uniqueness

Taken together, this realist ontology and this object-dependant epistemology have a remarkable implication: token physical objects can be *identified* and *re-identified* by cognizers, because objects fulfill sameness criteria which are traceable over time by cognitive means. A fundamental challenge is thus to determine *how* cognizers solve the uniqueness problem for object tracking, that is to represent and track token objects in a way which preserves the link with their identity (or uniqueness). Several analyses[3] have introduced an intuitive answer to this question, either as a useful metaphor (Perry, 2001a: 51-2) or as explanatory concept. The answer consists in the supposition that agents can store information about individuals in specialized, unique 'repertories' or 'files'. The two seminal ideas are that, in standard cases, (1) the uniqueness of the mental repertory ought to be related to the uniqueness of the target individuals in the world and that (2) one has to distinguish the creation and maintenance of a file from its purely descriptive content. These repertories are frequently called 'object files', and have been discussed in the theory of perception and language. Several notions of 'object file' remain to be distinguished, specifically in terms of the relation they bear to memory.

With respect to object perception, a number of authors have hypothesized the existence of a perceptual ability to manipulate (short term) *perceptual* object files (Kahneman et al., 1992; Pylyshyn, 2003). An *object file* in this sense is a mechanism that allows keeping track of an individual in a perceptual field. Kahneman and his colleagues (Kahneman & Treisman, 1984; Kahneman et al., 1992) have suggested that we store information about visually presented objects in *visual* object files. They introduced the notion in these terms:

> "Some time ago we proposed an account of object perception as the process of setting up and utilizing temporary 'episodic' representations of real world objects, which we call object files (Kahneman & Treisman, 1984). These are separate from the representations stored in a long-term recognition network, which are used in identifying and classifying objects." (Kahneman et al., 1992: 176)

These authors introduced this notion of object file along with the distinction between *temporary representations* of objects and *long-term recognition* of objects. This distinction is justified by Kahneman et al. (1992: 176) mostly by empirical evidence, including the primacy of objects in determining the allocation of attention (e.g., Baylis & Driver, 1993; Duncan, 1984; Egly, Driver, & Rafal, 1994; Pylyshyn, 2003: 181-200; Scholl, 2001). Attention to any one property of an object causes irrelevant properties of that object to be attended to, as in the Stroop effect (Stroop, 1935). Moreover, the division of attention between relevant attributes is facilitated if the attributes belong to the same object (Treisman, Kahneman, & Burkell, 1983). Kahneman et al. (1992) have also found an 'object-specific matching effect': the focusing of attention of a target object not only enhances the salience of all its current properties – it also selectively reactivates the recent history of that object.

In the same tradition, Pylyshyn (2001; 2003) has further suggested that these object files are 'linked to' individual distal items via an indexing mechanism called a 'visual index' (or 'FINST'). This analysis allows not only to account for the encoding of object-property information in a file, but also for its production and persistance. According to this kind of analysis, there are at least three classes of properties relevant to the understanding of the dynamics of tracking individual objects. First, *index-capture features/cues* are properties that cause the assignment of an index (and an object file to be opened). Second, *index-preservation features/cues* are properties that allow the indexed item to be tracked over time (and the file to persist over time). Third, *encoded features/cues* are properties that can be ascribed to the indexed item (information that can be stored in the file). The encoded features or cues are typically considered as *descriptions* that are suitable for the individual whose appearance has triggered the opening of the file; they are frequently said to be responsible for the visual awareness of the object. Taken together, these three kinds of properties characterize how 'object files' function. The *narrow* content of a file, i.e. information attached to the indexed individual, is defined by encoded features/cues. The dynamic anchoring of the tracking is defined by index-grabbing and index-preservation cues. The *broad* content of a file is defined by its existential dependency with respect to the causal properties of the external target object.

---

[3] For object files in thought and language, see e.g., Perry (1980; 2001a; 2001b), Bach (1987), Recanati (1993) ; for object files in vision, see e.g. Kahneman et al. (1984; 1992), Pylyshyn (2003), Palmer (1999), Saiki (2003). Interesting skeptical remarks can be found in Millikan (1997).

4 Extending the 'object file' notion to agent tracking

Human beings seem to be able to deal with two distinct classes of individuals. On the one hand, they are able to interact with and reason about *non-intentional* physical objects (e.g., Baillargeon, 2001; Spelke, 1990; Spelke et al., 1995). Tracking the observable behavior of material objects requires a detection or understanding of some of the regularities underlying physical phenomena. On the other hand, humans can also interact with and reason about *intentional agents*, i.e. individuals to whom we can ascribe *intentional states* such as beliefs, desires and intentions (e.g., Bloom, 2002, 2004; Dennett, 1971; Malle, Moses, & Baldwin, 2001). In order to account for the observable behavior of agents it seems that we need to add further principles, principles not reducible to those used in tracking physical objects, and which depend on intentional agency. This distinction between two domains in which different individuation mechanisms are at work, is what Bloom calls "commonsense dualism". According to his analysis of common sense (Bloom, 2004: 191), dualists have two ways of looking at the world: in terms of bodies and in terms of souls; a direct consequence of this dualism is the idea that bodies and souls are separate.

There is an interesting way in which singling out and tracking intentional agents might be explained by appealing to the notion of 'agent file' akin to the notion of 'object file' in vision. This approach has already been mentioned at least in philosophy. Perry (2001a: 51-2; 2001b: 123-46) suggests that information about agent identity can be stored in 'files' and that acquiring further perceptual information about an agent can lead to the opening, merging, splitting, or enriching a file. *Prima facie*, it seems reasonable to assume that in tracking agents humans might exploit the same taxonomy of properties that we have introduced for object tracking. We need to *pick out* a single intentional entity, to *keep track* of it across time and *ascribe* information to it. According to this analysis, one should enumerate *agent-indexing features/cues* that are properties that enable the anchoring in an intentional agent – and to open an 'agent file'; *agent-preservation features/cues* that are properties that allow the indexed agent to be tracked – and the file to persist over time; *encoded features/cues* that are properties that are described by symbols within the file and that are ascribed by the system to the tracked agent (information that is be stored in the agent file).

The conjecture that there exists such a dedicated mechanism for tracking an agent – an 'agent file' – raises the question of understanding whether and why such files are related to 'object files' that apply to tracking non-intentional objects. Here we are presented with at least three distinct options. According to an *independence view of agent and object tracking*, object files and agent files do not share any common feature and are dedicated to distinct domains. According to a *deflationary view of agent tracking*, there is only one type of mechanism that keeps track of and refers to either an object or an agent; agent files are thus reduced to object files. According to an *object-dependence view of agent tracking*, although there can be important differences between tracking physical objects and intentional agents, we track intentional agents by tracking objects. In the following sections, we will argue in favor of the third option, the object-dependence view.

5 The object-dependence view of agent tracking: Tracking agents *by* tracking physical objects

A number of arguments support the idea that tracking an agent requires tracking it on the basis of its status as a physical object because ordinary tracking of intentional agents seems to rely on a robust correlation between agents and bodies. One attractive argument runs as follows: given the usual connection between intentional agents and their bodies, on the one hand, and agents' bodies and objecthood properties, on the other, being able to perceptually track objecthood properties (i.e. physical objects/bodies) might provide a means of accessing and keeping track of intentional agents.[4] Somewhat more formally, one argument for the object-dependence view is thus as follows:

> *(1) Thesis of the co-instantiation*: A human agent *a* is co-instantiated with his/her human body *b*.
>
> *(2) Thesis of body tracking*: The tracking of a human body *b* requires tracking *b*'s object properties.
>
> *Conclusion*: Hence, the tracking of a human agent *a* requires the tracking of the object properties of his/her body *b*. (Note that the conclusion should hold not only for humans but also *mutatis mutandis* for primates

[4] Here, we do not discuss directly the specific cases of self-identification and personal identity (e.g., Parfit, 1971; Quinton, 1962; van Inwagen, 1997) – see section 6.1.

and other animal species.)

The thesis of the co-instantiation is the one most open to conflicting interpretations and ontological debates (see the next section). We shall first comment on the thesis of body tracking, which is far less subject to controversy. It seems simply to fall out of an examination of the common sense assumptions about the spatio-temporal characteristics of human bodies, in a way that does not conflict with common assumptions in cognitive science. For, as a biological organism, the body of a human being fulfills a number of objecthood criteria.[5] Such a body has *cohesive* parts (limbs, organs) that are functionally and hierarchically organized. It possesses a defined *shape* which is delimited by surfaces (skin) that are boundaries between internal parts and external phenomena or objects. It is mostly impenetrable in a weak sense: a human body cannot strictly coincide in space with another human body. Several adult human bodies can only be in contact or close contact but cannot fuse or overlap in space. Each body can be counted among groups of several individuals because it is a unit. In addition, phenomena which threaten the cohesion and unity of the body are threatening the very existence of that body as a living unit. This unity is linked to its uniqueness. One of the clues to this uniqueness is that, as with ordinary physical objects, each body follows a continuous and unique spatio-temporal path from birth to death. As a result, this body can be located within different frames of reference (egocentric, allocentric) and many of the inferences that one can make about a human body are based on an implicit knowledge of the previous characteristics. For instance, if one knows that Paul is in the aircraft number 743, one can infer the spatio-temporal location of Paul's body on the basis of the knowledge of that aircraft's allocentic position (this is a cognitive way of keeping track of a physical target object, in absence of present perception). In addition, the eventuality of the aircraft's crash has its deadly implications *because of* our persistent assumptions of all the previous characteristics (that is, the physics of the human body is known to be such that it *will be* destroyed by any event that would cause destruction of the aircraft).

Given that human bodies fulfill these spatio-temporal objecthood criteria, the premise of body tracking seems unproblematic. For object properties/cues are the main available target for object perceptual tracking and the main evidence for reasoning about the spatio-temporal paths of objects which are not currently perceived. In addition, this last assumption seems consistent with recent evidence on perceptual tracking systems. Consider first the case of visual tracking and visual attention. Many of the properties which are assumed to be used in order to open and maintain a visual object file (or to assign a FINST) are object properties in the sense that they fulfill objecthood criteria (Carey & Xu, 2001; Kahneman et al., 1992; Pylyshyn, 2003). In particular, they are related to *cohesion* and *spatio-temporal continuity*. For instance, in the standard MOT experiment (Pylyshyn & Storm, 1988), the only available feature to distinguish targets from non targets is the spatiotemporal path followed by each element. Although the system for visual tracking can deal with short periods of occlusion, tracking abilities are clearly impaired as soon as the target element infringes on the cohesion principle via transformations similar to "pouring" (vanMarle & Scholl, 2003) or exploding. It is also difficult to track particular parts of distinct objects (Scholl, Pylyshyn, & Feldman, 2001). These results are evidence that physical objects may be typical targets for visual tracking. Hence, given that human bodies fulfill the major objecthood criteria, human bodies may be typical targets for visual tracking. If the object-dependence view is true, this observation about visual/perceptual tracking may generalize beyond mere visual tracking to multimodal tracking, in a way which should be congruent with a variety of cognitive strategies for keeping track of individuals in the world.

The change blindness paradigm provides further evidence for the view that the perceptual tracking of intentional agents can be on occasion performed on the basis of the tracking of basic (spatio-temporal) objecthood cues. In one experiment by Simons & Levin (1998) an experimenter initiated a conversation with a pedestrian and during the interaction he was surreptitiously replaced by a different experimenter. Only half of the pedestrians detected the change. What this seems to suggest is the fact that the pedestrian might use a primitive mechanism such as a (visual) object file (or a FINST) to perceptually track the experimenter agent – a mechanism that uses spatio-temporal features for tracking instead of elaborate unique visual descriptions. In the case of detection failure, the basic mechanism might have been blind to the person change during the 'surreptitious replacement' because this event would be (incorrectly) interpreted as a temporary occlusion of the *same* visible object. This type of analysis is consistent with the deictic or interactive approaches to vision (Ballard, Hayhoe, Pook, & Rao, 1997; Churchland, Ramachandran, & Sejnowski, 1994; O'Regan & Noë, 2001; Pylyshyn, 2000), according to which situated vision is not a rich description of the visible scene but essentially involves direct relations to objects in the vicinity, which are probed or explored only when required.

[5] Cf. e.g. Strawson (1959: 87-116), van Inwagen (1990), Merricks (2001).

Moreover, while we think there is good reason to resist the stronger, deflationary view of tracking agents (section 6), it can seem that *at least* the object-dependence view has got to be correct. Recall Bloom's claim that humans are natural dualists who have two ways of looking at the world: in terms of bodies and in terms of souls and that a direct consequence of this dualism is the idea that bodies and souls are separate (Bloom, 2004: 191). Not only are they separate (ontologically distinct), but 'souls' are held to be immaterial, and hence not perceivable by the usual means. However, if this be the case then souls cannot be tracked directly; whereas, agents' bodies – which peoples' souls somehow 'inhabit' – *are* among 'the furniture of the world'. Hence, agents' bodies are the only public, tractable objects relating to agency even from the dualist viewpoint. (Cf. Wittgenstein: "The best picture of the human soul is the human body".) Given the assumption of common sense dualism of 'one body, one soul' (which seems to imply accepting the premise of the co-instantiation for the period of 'terrestrial life'), thus, 'commonsense dualism' lends itself to the object-dependence view.

It is essential to note, however, that essentially the same reasoning holds even once dualism is rejected. For while materialism about the mind entails that *intentional states* (agents' states such as beliefs or emotions), ultimately, are (or supervene upon) *physical states*, there is no suggestion among current materialist theories (simple behaviorists aside) that the intentional states are just the familiar, observable states of a person's body. Rather, it is usually said that what constitutes a given intentional state are certain states of a person's brain (e.g., Dennett, 1978a; Shoemaker, 1997), perhaps together with facts about those states' history, their relation to other such states, and so on. Materialists about the mind, of course, disagree over *just which* such physical states and facts constitute intentional states. But the important point is that none of the most promising and popular candidates are any more readily accessible by ordinary, unaided perceptual means than the immaterial 'souls' posited by the dualist. Once again, though, given that those states are states of an intentional agent whose body *is* directly observable, humans can track the former by tracking the latter.

6 Beyond the object-dependence view: the important differences between tracking *intentional agents* and *non intentional objects*

The evidence reviewed so far is compatible with what we called the object-dependence view of agent tracking. In the terminology of the object-file theory, the conclusion would be that object files may, on occasion, be put to use in tracking intentional agents; agent files are thus sometimes reduced to object files. However, there are reasons for thinking that, while we do track intentional agents by tracking their bodies, we should resist the stronger, deflationary view. For there appear to be important differences between tracking physical objects and tracking intentional agents.

6.1 The variety of ontological theories of personal identity

First, the distinction between object and agent tracking is required by the ontological theories of personal/self identity which are not consistent with the *thesis of the co-instantiation* expressed in section 5, which states that a human agent is co-instantiated with his/her material human body. The conflict would heighten if one replaces 'agent' by 'identical person' in the thesis of the co-instantiation. The latter version of the thesis is rejected by a number of metaphysics for which the criteria of personal identity are not reducible to objecthood/bodiliness criteria, but pertain to other criteria such as psychological continuity (e.g., Shoemaker, 1997) or the understanding specific to shared intentionality and social folkpsychology (e.g., Ames et al., 2001; Tomasello, Carpenter, Call, Behne, & Moll, in press). For instance, Shoemaker (1997) defends a functionalist analysis of personal identity which rests on conditions of psychological continuity that persist in spite of dramatic bodily transformations such as brain transplantations or teleportation. On the credit of an object-dependence view of personal identity, is worth noting however that the compatibility of Shoemaker's analysis with materialism remains subject to caution (van Inwagen, 1997). In addition, a number of metaphyscians defend ontological analyses of personal identity which seem consistent with the thesis of the co-instantiation for personal identity. For instance van Inwagen (1990) and Merricks (2001)[6] uphold that each person is a material object and persists over time because this person is identical to the biological organism he/she *is (strictly)*. In the present article, our general use of 'agent' instead of 'person' is an attempt to remain as neutral as

[6] For instance : '(…) there is exactly *one* thing where we truly believe there to be a human person and a human organism (and a human body). Obviously, this implies that the person is identical with the organism (is identical with the body).' Merricks (2001: 86).

possible with respect to the puzzles of personal identity. Our claim focuses on agency and can remain true with a variety on different ontologies of personal identity.

6.2 Argument related to means of direct perceptual anchoring into agency (agent indexing)

Apart from the previous ontological considerations, arguments can also be found with respect to an examination of the perceptual indexing of agents. Even though this leaves intact the thesis of body tracking (section 5), there are important differences in the properties that can trigger and maintain singular agent perceptions – i.e. the opening/indexing and maintenance of 'agent files' –, and not just in the descriptive information carried out by the singular perceptual representation. One argument supporting this claim is that specific mechanisms seem to anchor the mind on properties that are usually co-instantiated with agency, and that appear as *direct* means for perceptually keeping track of agents (i.e. physical objects that instantiate agency properties).

For instance, humans can efficiently detect and track biological motions which are specific to agents. A tradition dating from the studies by G. Johansson (1973; 1975) showed how moving organisms can be detected purely from motion information. Johansson hypothesized that people would be able to perceive the movement of the human body from just the motion of the body's *joints*. To test this hypothesis, he filmed an actor in the dark with small lights attached to his joints (ankles, knees, hips, shoulders, elbows, and wrists) so that nothing was visible except the lights. When the actor was seated motionless in a chair, observers perceived a meaningless configuration of points, rather like a constellation of stars; nobody perceived the lights to be connected to a human agent. But within fractions of a second after the actor began to move, first standing up and then walking, he was immediately and unmistakably perceived as a human agent in motion. In other further studies, other researchers have found that observers are able to discriminate between male and female walkers who have lights placed just on their ankles, knees, and hips (Cutting, Proffitt, & Kozlowski, 1978) – see also Troje (2002) and the simulation of this phenomenon at http://www.journalofvision.org/2/5/2/genderclass.html.

Also, there is evidence that certain *types of motion* lead to the attribution of specific types of intentional states (e.g. Gergely, Nadasdy, Czibra, & Biro, 1995; Heider & Simmel, 1944; Jacob & Jeannerod, 2003: pp. 222-6). Heider and Simmel (1944) showed human adults an animation on a screen involving three geometric objects: a small circle, a small triangle and a large triangle moving in the vicinity of a very large non-moving square. Unlike the perception of static display of the three geometric objects, the perception of the kinetic structure of the patterns of motion of the objects conveyed psychological and even social information about the objects. Subjects used highly intentional verbs to describe the behavior of the triangles such as they 'chased', 'attacked', 'caressed' or 'comforted' the circle. In the same tradition, in reply to Atran (1998), Todd & López (1998: p. 592) describe experiments similar to the famous previous ones: 'We have explored a simple visual cue-based algorithm for judging intention from motion in just such instances (Blythe, Miller, & Todd, 1996). We had participants generate motions of two moving "bugs" on a computer screen, corresponding to simple intentional categories including pursuit, evasion, fighting, courtship, and play. Other participants were later able to categorize the intentions of the "bugs" with high accuracy from their trajectories alone. This study supports the notion that animate intention can be determined using only a few simple spatiotemporal cues (which include, from trajectory analysis, relative heading, relative distance, relative velocity, and vorticity or "loopiness"). Knowing the intention (as opposed to the general intentionality that Atran mentions) of another organism can trigger the appropriate domain-specific mechanism for response, including species-level categorization and recall of relevant traits.'

Moreover, in the visual domain, there is evidence that the visual system does not rely on the same resources for recognizing non-face objects as opposed to faces (e.g. Grill-Spector, Knouf, & Kanwisher, 2004). This might suggest that the presence of a face – an agent specific cue – can contribute to a specific file indexing, an agent-based indexing. This could be the case not only in vision, but also in auditory perception and cross-modal integration. It seems plausible, for instance, that noises that bear the specific signature of human or animal behavior trigger the opening of an agent file – i.e. we can detect and track a person just by hearing a sound that bears the specific signature of a human body (voice, footsteps). This is the case when one is detecting a sound related to the phonological part of the body (e.g. Handel, 1995).

6.3 Argument from parsimony for the direct access to agency

The empirical data drawn from the study of early motor and sensory skills seem to support the idea that there may even be two distinct kinds of files. What would be the use of such a distinction from a computational perspective – as introduced by Ballard et al. (1997) and Pylyshyn (2001)? It is often necessary to direct the focus of attention

specifically toward agents, as opposed to mere non intentional objects, for instance in order to achieve collaborative actions (e.g. shared intentionality in word learning or collective sports), or to evaluate judgments involving intentional relations. If the deflation view is true, the only difference between object and agent files would be in terms of the descriptive content they encode. This would imply that there would have to be a (descriptive- or content-driven) search for intentional properties in object files, which seems computationally costly. As opposed to this approach, one can speculate that the addressing mechanism of each file was typed *as object* or *as agent*; however, such a search would be made easier from a computational perspective. For instance, one might speculate that files could be sorted by a non conceptual 'tag' signaling directly that it is a '(mere) object file' or an 'agent file'; such a tag would avoid the need to access the descriptive content of a file so as to target one of the two kinds of entity for any attentional or motor routine (e.g. routines such as 'escaping from a predator', 'seeking for help' or 'searching for a team member').

6.4 Rationality and agency

Evidence for/against rationality is also relevant for tracking intentional agents, but not objects (which, since they are not intentional agents at all, are neither rational nor irrational). Thus, e.g., if someone consistently violates obvious norms of reasoning (e.g., *modus ponens*), it becomes difficult to even make sense of them as intentional agents [e.g., Dennett (1969; 1971), Davidson (1980; 1984)]; though of course, it may be not difficult at all to continue to track their bodies. Similarly, when someone consistently exhibits markedly different patterns of behavior/reasoning in different domains (cf. discussions of 'compartmentalized thinking', and/or 'false consciousness'), even if these differences do not correlate to marked differences in features of their body, we sometimes find it necessary to suppose, in effect, that there are two agents there (less dramatically, that markedly different psychological states and traits characterize the agent in the two domains) even though there may be seamless continuity in his/her single body. This assumption is frequently made about people supposed to be 'possessed' as depicted in ghost movies or stories; it is manifested in a pathological way in syndromes such as the Capgras syndrome where the patient believes that his/her near ones have been replaced by substitutes occupying the same body. In short, considerations of a conspecific's (ir-)rationality is a criterion relevant to the opening, maintenance, and updating of agent files, but it is not relevant at all to (non intentional) object files; and perceived discontinuities or dramatic differences in an agent's intentional states themselves can lead us to 'split' or seriously amend the relevant agent file, even though the corresponding object file (used in tracking his/her body) remains intact, and does not undergo comparable changes in its contents.

7 Conclusions

Assuming a realist ontology of the material world and an object-dependent epistemology, we have examined the possibility of tracking not only physical objects but intentional agents. Our contention, referred to as the object-dependence view, is that tracking intentional agents requires the same kind of processes used in tracking physical objects. Furthermore, we postulated, in a manner consistent with the findings of experiments conducted on visual perception, that it is possible to draw a parallel between the classes of properties that characterize and produce agent and object files (indexing, preservation and encoded properties or cues). Given the similar – though *not*, we have argued, identical – anchoring, content and architecture of (non intentional) object files and agent files, the exact relation between them remains an open question which we have only begun to explore. However, at least the following seems plausible: (1) we perceptually and cognitively keep track of intentional agents, (2) object files and agent files are distinguishable in important ways, but nevertheless (3) agent tracking exploits the resources of object tracking by anchoring the agent file on perceptible features of agents' bodies (thesis of body tracking).

8 Possible directions for future research

Our object-dependence view suggests directions for further research. A number of classical experiments have been carried out on object perception and multiple-object tracking in vision, in various conditions including conditions of target occlusions. It would be interesting to see what the results would be in analogous experiments in the case of visual multiple-agent tracking. What sorts of changes (in motion? behavior? intention?) are/are not relevant to subjects for maintaining the same agent file rather than opening a new one? Can one produce in an experimental display cases where persons-as-bodies and persons-as-agents 'come apart' (as when 'the soul rises to heaven', or 'body-switching' cases). Presently, the need first lies not in answering these questions but in raising them.

However, there is plenty of anecdotal evidence we could cite. For example, we have no difficulty in conceiving of

'body-switching' – an agent simply occupying a different body. For a 'fictional-indexical' descriptions of this, see Dennett (1978b); for a metaphysical conception, see the examples given by Parfit (1971) and Shoemaker (1997). This is not incompatible with our view. In fact, it might support it because the object-dependence can be exquisitely discovered in the ways one conceives of or depicts the 'traveling souls'. For, depictions of souls leaving a particular body often take the form of *ghostly* bodies leaving the body of flesh-and-blood. This is not surprising, for how else could souls be depicted – or, we ask, how else could they be *tracked*? Is it not by depicting human bodies that painters have been representing the souls of 'The Blessed' in Paradise and of 'The Damned' in Hell? See for instance 'The Last Judgement' (1431) by Fra Angelico (http://www.abcgallery.com/A/angelico/angelico39.html, and the details of the blessed (http://www.abcgallery.com/A/angelico/angelico41.html) and the damned (http://www.abcgallery.com/A/angelico/angelico42.html) in the picture.

We *do* have some trouble making sense of cases of fission/fusion (where one agent becomes two, or two become one – a good illustration for this could be found in the view of Lewis (1976). Here too, the cases might support our position: fission/fusion are puzzling, not merely because they involve splitting/joining minds, but because our usual means of tracking agents (via their bodies) is *frustrated* – the one-body/one-soul correlation is disrupted.

Acknowledgments: We would like to thank for their comments on earlier versions of this paper Erika Marchetto, Nivedita Gangopadhyay and the members of the 'agent tracking group' of CNRS Thematic School 'Reference to objects' (http://www.objectreference.org/). Part of this research has been founded by the Enactive Network.

9 References

Ames, D. R., Knowles, E. D., Morris, M. W., Kalish, C. W., Rosati, A. D., & Gopnik, A. (2001). The social folk theorist: Insights from social and cultural psychology on the contents and contexts of folk theorizing. In B. F. Malle & L. J. Moses & D. A. Baldwin (Eds.), *Intentions and Intentionality* (pp. 307-329). Cambridge, MA: MIT Press.

Atran, S. (1998). Folk biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral and Brain Sciences, 21*, 547-609.

Bach, K. (1987). *Though and Reference*. Oxford: Clarendon Press.

Baillargeon, R. (2001). Infants' Physical Knowledge: Of Acquired Expectations and Core Principles. In E. Dupoux (Ed.), *Language, Brain, and Cognitive Development* (pp. 341-361). Cambridge: MIT Press.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20*(4), 723-767.

Baylis, G. C., & Driver, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location. *Journal of Experimental Psychology : Human Perception and Performance, 19*, 451-470.

Bloom, P. (2002). Mindreading, communication and the learning of names for things. *Mind & Language, 17*(1&2), 37-54.

Bloom, P. (2004). *Descartes' Baby*. New York: Basic Books.

Blythe, P. W., Miller, G. F., & Todd, P. M. (1996). Human simulation of adaptive behavior: Interactive studies of pursuit, evasion, courtship, fighting, and play. In P. Maes & M. J. Mataric & J.-A. Meyer & J. Pollack & S. W. Wilson (Eds.), *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation and Adaptive Behavior*. Cambridge, MA: MIT Press.

Campbell, J. (1994). *Past, Space and Self*. Cambrige, MA: MIT Press.

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

Carey, S., & Xu, F. (2001). Infant's knowledge of objects: beyond object files and object tracking. *Cognition, 80*, 179-213.

Churchland, P. S., Ramachandran, V. S., & Sejnowski, T. J. (1994). A critique of pure vision. In C. Koch & J. L. Davis (Eds.), *Large Scale Neuronal Theories of the Brain* (pp. 23-60). Cambridge, MA: MIT Press.

Clark, A. (2000). *A Theory of Sentience*. Oxford: Clarendon Press.

Cutting, J. E., Proffitt, D. R., & Kozlowski, L. T. (1978). A biomechanical invariant for gait perception. *Journal of Experimental Psychology: Human Perception & Performance, 4*(3), 357-372.

Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Clarendon Press, Oxford University Press.

Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

Dennett, D. C. (1969). *Content and Consciousness*. London: Routledge and Kegan Paul.

Dennett, D. C. (1971). Intentional systems. *The Journal of Philosophy, 68*(4), 87-106.

Dennett, D. C. (1978a). *Brainstorms*. Montgomery: Bradford Books.

Dennett, D. C. (1978b). Where am I?, *Brainstorms* (pp. 311-323). Montgomery: Bradford Books.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology : General, 119*, 501-517.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology : General, 123*, 161-177.

Evans, G. (1981). Understanding demonstratives. In H. Parret & J. Bouveresse (Eds.), *Meaning and Understanding* (pp. 280-303). Berlin: De Gruyter.

Evans, G. (1982). *The Varieties of Reference*. Oxford: Oxford University Press.

Gergely, G., Nadasdy, Z., Czibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*, 361-370.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience, 7*(5), 555-562.

Handel, S. (1995). Timbre perception and auditory object identification. In B. C. J. Moore (Ed.), *Hearing* (pp. 425-461). San Diego, CA: Academic Press.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Jounal of Psychology, 57*, 243-259.

Jacob, P., & Jeannerod, M. (2003). *Ways of Seeing, The Scope and Limits of Visual Cognition*. Oxford: Oxford University Press.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*(210-211).

Johansson, G. (1975). Visual motion perception. *Scientific American, 232*(6), 76-88.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of Attention* (pp. 29-62). Orlando: Academic Press.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*(2), 175-219.

Kripke, S. (1980). *Naming and Necessity*. Oxford: Basil Blackwell.

Lewis, D. (1976). Survival and identity. In A. Rorty (Ed.), *The Identity of Persons*. Berkeley: University of California Press.

Malle, B. F., Moses, L. J., & Baldwin, D. A. (2001). Introduction: The significance of intentionality. In B. F. Malle & L. J. Moses & D. A. Baldwin (Eds.), *Intentions and Intentionality* (pp. 1-24). Cambridge, MA: MIT Press.

McDowell, J. (1984). De Re Senses. *Philosophical Quarterly, 34*(136), 283-294.

McDowell, J. (1990). Peacocke and Evans on demonstrative content. *Mind, 99*(394), 255-266.

McDowell, J. (1998). *Meaning, Knowledge, and Reality*. Cambridge, MA: Harvard University Press.

Merricks, T. (2001). *Objects and Persons*. Oxford: Oxford University Press.

Millikan, R. G. (1997). Images of identity: In search of modes of presentation. *Mind, 106*(423).

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24*(5), 939-1031.

Palmer, S. E. (1999). *Vision Science, Photon to Phenomenology*. Cambridge, MA: MIT Press.

Parfit, D. (1971). Personal identity. *The Philosophical Review, 80*(1), 3-27.

Peacocke, C. (1991). Demonstrative content: a reply to John McDowell. *Mind, 100*, 123-133.

Perry, J. (1980). A problem about continued belief. *Pacific Philosophical Quarterly, 61*, 317-332.

Perry, J. (2001a). *Knowledge, Possibility, and Consciousness*. Cambridge, MA: MIT Press.

Perry, J. (2001b). *Reference and Reflexivity*. Stanford: CSLI Publications.

Pylyshyn, Z. W. (2000). Situating vision in the world. *Trends in Cognitive Sciences, 4*(5), 197-207.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*, 127-158.

Pylyshyn, Z. W. (2003). *Seeing and Visualizing: It's Not What You Think*. Cambridge, MA: MIT Press.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independents targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*(3), 179-197.

Quinton, A. (1962). The soul. *The Journal of Philosophy, 59*(15), 393-409.

Recanati, F. (1993). *Direct Reference: From Language to Thought*. Oxford: Blackwell Publishers.

Saiki, J. (2003). Feature binding in object-file representations of multiple moving items. *Jounal of Vision, 3*, 6-21.

Scholl, B. J. (2001). Objects and attention: the state of the art. *Cognition, 80*, 1-46.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition, 80*, 159-177.

Shoemaker, S. (1997). Self and substance. *Noûs, 31*(Supplement: Philosophical Perspectives, 11, Mind, Causation, and World), 283-304.

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review, 5*(4), 644-649.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14*, 29-56.

Spelke, E. S., Gutheil, G., & Van de Walle, G. (1995). The development of object perception. In S. M. Kosslyn & D. N. Osherson (Eds.), *Visual Cognition: An Invitation to Cognitive Science, Second Edition* (pp. 297-330). Cambridge, MA: MIT Press.

Strawson, P. F. (1959). *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen.

Strawson, P. F. (1997). *Entity and Identity and Other Essays*. Oxford: Clarendon Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Todd, P. M., & López, A. (1998). Pulling the trigger on the living kind module. *Behavioral and Brain Sciences, 21*(4), 592.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (in press). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*.

Treisman, A., Kahneman, D., & Burkell, J. (1983). Perceptual objects and the costs of filtering. *Perception & Psychophysics, 33*, 527-532.

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Jounal of Vision, 2*, 371-387.

van Inwagen, P. (1990). *Material Beings*. Ithaca, London: Cornell University Press.

van Inwagen, P. (1997). Materialism and the psychological-continuity account of personal identity. *Noûs, 31*(Supplement: Philosophical Perspectives, 11, Mind, Causation, and World), 305-319.

vanMarle, K., & Scholl, B. J. (2003). Attentive tracking of objects versus substances. *Psychological Science, 14*(5), 498-504.