Auditory and cross-modal attention for the cognitive access to objects

Nicolas J. Bullot

Auditory and cross-modal attention for the cognitive access to objects

Nicolas J. Bullot,

Institut Jean Nicod

CNRS EHESS ENS

1 bis avenue de Lowendal

75007 Paris

France

Auditory and cross-modal attention for the cognitive access to objects

0 Introduction

This article/presentation aims at studying how we track and identify objects on the basis of multimodal perception. It belongs to 'procedural' theories according to which demonstrative identification depends on using *procedures of perceptual attention* (e.g., Campbell, 2002; Clark, 2000; Evans, 1982; Pylyshyn, 2003; Ullman, 1984). In contrast to prevalent views according to which demonstrative identification is primarily based on the orienting of *visual* attention to the target object itself (Campbell, 2002: 115-16), I shall investigate an alternative Crossmodal View. According to the Crossmodal View, demonstrative identification depends more fundamentally on *crossmodal attention*. I shall present an argument maintaining namely that perceivers routinely use the coordinating ability of crossmodal attention to retrieve the *continuity* and *uniqueness* of the spatiotemporal path of the target object of their identification acts. The analysis will focus on examples of crossmodal links between audition and vision.

1 The theoretical background: outline of the deictic theory

I will start with a brief and schematic outline of a theoretical framework that can be called the *deictic theory*, [1] which rests on the link between direct reference and attention. One of the possible versions of the deictic theory relies on the following assumptions:

First, there is a requirement of focal attention (in perception) for performing deictic/demonstrative reference to one object during exploration of a spatial layout. Perceptual

---

[1] The deictic theory deals with perception conceived as representation of, or access to, a *unique* distal individual – i.e., a singular, or episodic, representation or thought – throughout a unique *episode*. The syncretism of the naming 'deictic theory' intends to express broad principles of potential agreement. This syncretism aims at narrowing the gap between works in philosophy of deictic/singular reference (Burks, 1949; Campbell, 2002; Kaplan, 1989b; Peirce, 1931-35; Recanati, 1993; Russell, 1984; Woodruff Smith, 1989) and psychology of attention (Driver & Baylis, 1998; Kahneman & Treisman, 1984; Kahneman, Treisman, & Gibbs, 1992; Pylyshyn, 2001, 2003; Scholl, 2001; Treisman, 1969). A number of works have (put) emphasised on the relationship between attention and direct reference. Regardless of their diverging views of what attention is precisely, their works exhibit a common interest for relating attention and perceptual deictic identification. These works include early contributions by Peirce (Burks, 1949; Peirce, 1931-35), James (1890a) or Russell (1911; 1956). Among recent contributions, John Campbell (2002), Austen Clark (2000) and Zenon Pylyshyn (2003) defend distinct versions of this deictic theory. The latter authors have introduced explicit considerations about *attention* which were almost absent from the seminal contributions of Evans (1982), Kaplan (1989a; 1989b) or Perry (1977; 1979).

deictic reference to a particular physical object (located within the explored spatial layout) requires the cognitive agent (i) to position the target object within the reach of at least one sensory field (e.g., visual, tactual or auditory) and (ii) to select this object by focal attention.[2] Both conditions are necessary for reference to be genuinely based on perception; the second one is frequently conceived of as a sufficient condition.

Secondly, this attentional requirement has also to be fulfilled for performing *deixis*-based communication.[3]

Thirdly, this attention requirement rests on a situated *cognitive access*, for attentional systems secure the cognitive access to each target object or cue. There is thus a fundamental *accessibility problem* studied by competing theories about attention: how does the attentive mind get cognitive access to the target objects or cues?

2 The puzzle of visual and crossmodal identification

A number of traditions are studying how attention relates to identification and demonstrative reference. The role of *visual* attention is perhaps the most intensely studied topic (e.g., Campbell, 2002; Pylyshyn, 2003; Scholl, 2001), namely because vision informs about the spatial layout of objects and scenes in a way that is valuable to identifying things in the world. Although vision might be the 'primary' sense for recovering the spatial layout of objects, there is no justification to conceive of this sense as *the* exclusive system for identifying objects. Perceivers can scrutinize, recognize and perhaps identify objects on the basis of a non-visual sensory systems such as audition (e.g., Bregman, 1990; Handel, 1995; McAdams & Bigand, 1993; Wightman & Jenison, 1995) or haptic perception.[4] However, the question persists as to whether identification can be performed within a single 'pure' sensory system or not. For instance, can one identify and re-identify one particular object on the basis of mere auditory experience? Can one entertain genuine singular thoughts based on mere auditory experience? These questions are puzzles for philosophy and psychology.

---

[2] See for instance Campbell (2002), Mack & Rock (1998), Pylyshyn (2001; 2003), Yantis (1998).

[3] In communication based on perception, demonstrative referring to an object for someone else's benefit involves directing his or her focal attention to the target object. See for instance Baldwin (1993), Butterworth & Grover (1990), Campbell (2002), Mahle et al. (2001), Tomasello (1995). A number of bodily activities contribute to the achievement of this goal, such as carrying out pointing gestures with the index, head or gaze.

[4] One frequently describes everyday auditory experience in terms of perceiving and identifying resonating objects within a spatio-temporal layout – see for instance, Metzger (1953: 59-60), quoted in Plomp (2002: 10), and also Wightman & Jenison (1995) and Handel (1995).

A number of authors seem to believe that, when compared namely to auditory attention, visual experience exhibits an epistemic privilege for demonstrative identification. This is the *claim of the epistemic primacy of vision for identification*. [5] According to this view, vision provides a direct access to the categorical ground of the target object "itself". In other words, vision offers a privileged access to the causal and spatiotemporal properties of the object. For instance, John Campbell (2002) seems to endorse this claim in his own analysis of the contrast between *pure audition* and *visual experience* (Campbell, 2002: 115-16). His analysis is based on the following example. Suppose that you are just able to hear sounds coming from the house next door (for a rather long period). On the basis of 'pure' auditory experience, you can formulate hypotheses about the functional role of the sound sources within that particular house (e.g., 'This sound might come from a baby who is crying.'). [6] According to Campbell's analysis:

> Suppose now that the day finally arrives when you do get a look inside the house. What does this add to your knowledge? (…) The contrast between the knowledge you have now, on the basis of a look at the objects, and the knowledge you had before of the existence of objects with particular functional roles, is that when you see the thing, you are confronted by the individual substance itself. On seeing it, you no longer have knowledge of the object merely as the postulated occupant of a particular functional role. Your experience of the object, when you see it, provides you with the knowledge of the categorical grounds of the collections of dispositions you earlier postulated. (Campbell, 2002: 115-16)

According to Campbell, *visual* experience of the object can confront you with the categorical basis of the dispositional relations in which the object may stand to other things. As opposed to vision, 'pure' auditory experience cannot ground our knowledge on the things themselves, because auditory experience provides only information for general thoughts (instead of singular/particular thoughts) based on the postulation of particular functional roles.

Although vision seems indeed crucial for navigating and identifying things in space, Campbell (at least in this text) did not consider an alternative view, which contends that object identification is primarily a multimodal and crossmodal phenomenon, in such a way

---

[5] See also Strawson (1959).

[6] "Suppose that you live in one terraced row of houses, and you sometimes hear noises from the house next door. Being of an enquiring turn of mind, you formulate hypotheses about what objects are to be found next door. There's a couple, you conjecture, a man and a woman, though you never see them directly – there are two cars outside, you occasionally hear noises coming simultaneously from different parts of the house next door, and sometimes there are raised voices. (…) You could say: There are object x_1, x_2, …, x_n, which stand to one another in the following relations …, and which stand in the following relations to the audible phenomena … . You could, in effect, give a functional characterization of all the particular objects you postulate as being next door. And you might test and confirm your hypotheses over a long period, without ever catching sight of those things."

that no single 'pure' sensory modality (not even vision) is completely able to provide the required information to identify objects.[7] The latter approach can be named the 'crossmodal view' of demonstrative identification. It is clearly in conflict with the claim of the epistemic primacy of vision.

3 Hypothesis of crossmodal attention for identification

With respect to multimodal perception, procedural theories of attention lead to the following hypothesis of Crossmodal Attention for Identification ('CAI' henceforth):

> *Crossmodal Attention for Identification (CAI)*: Crossmodal attention is a necessary – and possibly sufficient – condition of the ability to keep track of, and identify particular distal objects.

Two complicated concepts are central to this hypothesis: *crossmodal attention* and *identification*. I shall discuss both of them. In the context of experimental literature, the concept of *crossmodal attention* is used to refer to *capacities* and *effects* involved in the process of coordinating – or 'matching' – the information picked up by multiple perceptual modalities (e.g., Driver & Spence, 1998: 254-55, 259-60). For instance, during the multimodal perception of one unique (audible and visible) speaker, the concept of 'crossmodal attention' is frequently used to refer to the capacity to coordinate the information picked up by audition (speech perception) and vision (lip-reading). *Prima facie*, crossmodal attention should have the function to resolve the problem of how *several* distinct bits of modal[8] information relate to one *single* element (event or object). The problem with this notion is that there is no available consensus on the mechanisms responsible for crossmodal effects and on how they relate to attention (exogenous or endogenous). In particular, the literature oscillates between a notion of crossmodal attention that relates to *nonconceptual skills* that causally contribute to cognitive prioritized selection and *conceptual capacities* that resolve problems based on discrepancies between distinct pieces of modal information.

According to the crossmodal view, crossmodal attention has a fundamental role in the identification of particular objects.[9] The concept of *identification* here stands in need of

---

[7] In other words, I would like to argue for the view according to which our relationship to the properties of the categorical basis of each target object of demonstrative identification takes place in the background of multimodal perception and tracking.

[8] That is to say, picked up with distinct modalities.

[9] J. Campbell (2002) makes the case for a similar analysis – expressed within a Fregean framework – but focuses rather on conscious attention. He (Campbell, 2002: 88) asserts that: "(…) different ways of consciously attending

clarification, for it can refer to a more or less sophisticated ability (Clark, 2000: 144-151; Dretske, 1995: 334; Millikan, 1984: 239-256). Under one of the weakest interpretations, to 'identify' in perception amounts to detecting or recognizing only *types* of properties, instead of *tokens* of individuals (i.e., token distal objects). Under one of the strongest interpretations, 'to identify' means that the agent performing the identification has the knowledge to individuate and re-identify one *token* individual object from all other tokens of the same type. For instance, Evans' theory of demonstrative identification (Evans, 1982) refers to a strong interpretation.[10] Evans' analysis emphasizes that demonstrative identification requires the capacity to *locate the target object* in an egocentric space (Evans, 1982 : 178; 1985). More generally, in an Evansian account, the demonstrative access to the target *uniqueness* is dependant on the capacity to keep track in perception of the *unique* spatio-temporal path of this particular target.[11]

This said, according to a procedural approach to attention, *ways of identifying* things are grounded into *ways of attending* to things. In accordance with this idea, the following argument may provide ground for the crossmodal view:

---

to a single object can cause and justify the use of different information-processing procedures to verify, and to find the implications for action of, different demonstrative propositions involving reference to the same object. To characterize the sense of a demonstrative, then, we need to know how to characterize the content of conscious attention to the object (…)".

[10] It is linked with "Russell's principle" about singular thought – according to which "the subject must have a capacity to distinguish the object of his judgment from all other things" (Evans, 1982: 89). This conception of the demonstrative identification of physical objects is particularly strong, since it requires the fulfilment of three criteria. First, "its previous and present deliverance**s** provide the subject with his governing conception of the object" (Evans, 1982: 174). Second, "the subject remains 'in contact' with the object, and is thus (unmediatedly) disposed to alter his governing conception in response to certain future information received from the object" (Evans, 1982: 174). Third, "the subject is able, upon the basis of the link, to *locate* the object in egocentric space, and thereby in objective space" (Evans, 1982: 174).

[11] Nonetheless, in the orthodox view advocated by Evans (1982), the perceptual tracking of the spatio-temporal path has to be completed by *conceptual* identification with a kind concept (a sortal) so as to lead to what he calls the "fundamental idea" of one particular object (Evans, 1982 : 178).

> *Proposition 1*: Crossmodal attention to one target object *x* – or a cue related to *x* – is a necessary condition required in active perception:
>
> A. In order to get perceptual access to *x*'s properties – via *x*'s continuous tracking by routines of distinct perceptual systems;
>
> B. In order to resolve epistemic identification queries about *x*'s accessible properties – via reasoning about, and conceptually keeping track of, some of *x*'s properties.
>
> *Proposition 2*: Resolving epistemic identification queries about the perceptually accessible object *x* is a necessary condition of the ability to keep track of and identify this object *x* within our spatiotemporal conceptual scheme[12].

From the above, we could draw the following conclusion:

> Hence, crossmodal attention to *x* is a fundamental condition of the ability to keep track of and identify this object *x* (within our perceptual fields and within our spatiotemporal conceptual scheme). (*Crossmodal Attention for Identification, CAI)*

According to the foremost idea of this argument, identification depends on the spatiotemporal keeping track of the target, which in turn depends on multimodal and crossmodal coordination. The distinction between proposition 1A and 1B suggests the possibility of two general kinds of crossmodal coordination. The first type is coordination for tracking the target across one's sensory fields. The second is coordination on the basis of conceptual queries and reflective thinking about the information provided by distinct sensory modalities.

If this argument be valid, Campbell's interpretation of the example – and the claim for the epistemic primacy of vision – should be at least revised for the access to the categorical basis of the object's dispositions is here primarily crossmodal. As a result, the relational character of demonstrative thought should be primarily grounded in crossmodal experience. (In addition, this might be consistent with the other view adopted by Campbell when he discusses crossmodal binding (e.g., Campbell, 2002: 126-31).)

Proposition 2 bears on the dependency of the conceptual ability to locate things in a spatiotemporal frame of reference, on the capacity to resolve identification queries about the perceptually accessible object. Epistemic identification culminates in demonstrative identification since it amounts to performing identification procedures ('epistemic queries') that can lead to observational judgments or beliefs. The idea is that one can get discriminative knowledge about physical objects, and that the conceptual knowledge one gets through perception helps in updating and locating the target within our conceptual scheme. This view is consistent with a number of accounts of *epistemic* perception (Dretske, 1979, 1995; Evans,

---

[12] One assumes here a notion of spatio-temporal scheme akin to Strawson's in *Individuals* (Strawson, 1959).

1982; Millikan, 1984; Strawson, 1959) concerning the perceptual picking up of information required for identifying the target. One can thus assume (2) as a rather standard framework in the epistemology of perceptual identification.

I would like to focus here on the first proposition (esp. so as to study how parts A and B of this argument can be articulated).

4 Accessibility routines and spatio-temporal continuity: the examples of crossmodal capture and crossmodal enrichment

Part A of proposition (1) concerns the nonconceptual anchoring capacities (or routines) involved in cognitive accessibility, which may presumably be fulfilled either by non representational mechanisms or by nonconceptual content. There are arguments for the necessity of these routines for visual attention.[13] Nonetheless, a number of distinct arguments support also a specific role for *nonconceptual crossmodal* attention. I shall discuss some examples that relate to audio-visual interactions.

Some arguments for 1A relate to the accessibility based on the phenomenon of exogenous attentional capture. *The general reason is that capture seems to entail crossmodal upshots, which contribute to the continuous perceptual tracking of distal elements across distinct sensory fields*. [14]

The notion of *exogenous capture* refers to an apparent truncation of a voluntary search due to the presence of a distinctive attractor element which "pulls" attention to a specified location or object. [15] Attractor elements may be for instance events with abrupt onsets. Capture relates thus to an involuntary access to (unexpected) events or objects that deserve attention due to

---

[13] For instance, Campbell (2002: 61-83) dismisses the view according to which visual identification can be explained by the mere use of a sortal concept. He refers to conscious attention as a nonconceptual capacity that can obtain visual binding and cognitive access to visible objects. Emphasizing a similar need for pre-conceptual access, Pylyshyn (2003: 200-79) hypothesizes non-representational mechanisms – visual indexes – that allow performing visual tracking.

[14] It has long been described by phenomenological analyses that attention can undergo involuntary shifts (see Hatfield, 1998: 10, for an historical overview). This has led to the distinction between (i) automatic or reflex and (ii) voluntary attention within various lexical idioms. For instance, James (1890b: 416-17) distinguishes between 'passive' and 'reflex' attention on one hand and 'active' and 'voluntary' attention on the other. At the end of the nineteenth century, Wundt (1897: 217-18) or Titchener (1899) mention also a related distinction. Experimental cognitive sciences brings also this distinction into play, but use the phrases 'exogenous attention' and 'endogenous attention' respectively (e.g., Driver & Spence, 1998; Jones, 2001; Spence, 2001). We will keep the latter for the present discussion.

[15] Controversies remain in psychology over the specific routines responsible for capture, and namely about there possible automaticity (Jones, 2001; Spence, 2001). Central to the debate over capture is a concern over whether certain stimulus properties, intrinsic to the distal attractor element (such as abrupt onsets), are responsible for the attentional shift. These controversies do not affect the general argument we are dealing with in the text.

salient features. [16] What is relevant for our discussion is that exogenous capture is a crossmodal phenomenon, since attentional capture *in one* perceptual modality facilitates *overt* and *covert* access *in other modalities access* (to the properties of the attractor element – object, cue or event).

First, consider the case of publicly observable bodily movements, the overt case. Consider a distal event constituted by the fall from a table of a heavy object, such as a book or a glass, and its collision with a parquet floor. The collision will cause an abrupt impact sound that operates as an alerting signal for surrounding perceivers. Typically, the acoustic and auditory event will elicit (overt) saccadic eye movements and bodily orientation toward the source of the sound (Driver & Spence, 1998; Gibson, 1966: 75). Hence, this event has usually typical *overt crossmodal* consequences for each perceiver in its vicinity. In particular, the auditory event triggers an *audiovisual perceptual tracking* of the sound source which may be followed by visuo-haptic tracking of the same (e.g., if the perceiver searches, reaches and grasps the object).

Second, this description seems underpinned by experimental findings about the covert outcomes of attention capture. For instance, experimental works with the 'orthogonal cueing' paradigm studied by Driver & Spence (1998) and colleagues indicate that multimodal priming and crossmodal links in exogenous spatial attention seem to occur even before any overt bodily motion, in a *covert* manner. Spatially non-predictive cue in one modality can attract covert attention towards its location in other sensory fields, not solely within the cued modality (Driver & Spence, 1998: 255). For instance, abrupt sounds seem to 'attract' visual and tactile attention, not merely auditory attention. If this be true then it suggests that the capture of auditory attention in a given region of space facilitates the detection *within the same particular region of space* of events/objects by other sensory modalities (Driver & Spence, 1998).

Hence, attentional capture seems to obtain (involuntary) cognitive spatial access to salient objects or cues *within one single multi-modally tracked location*. In addition, since the properties of each physical object are usually co-localized, this kind of crossmodal capacity is likely to facilitate the continuous tracking of the target object across distinct sensory fields.

---

[16] Exogenous attention is frequently said to be automatic, but the contrast between voluntary attention and automaticity remains a controversial matter (Kahneman & Treisman, 1984; Logan & Compton, 1998; Shiffrin, 1997).

(Another similar argument for the crossmodal nature of basic access-routines derives from speech perception with endogenous attention.[17])

It is appealing to conceive of this *overt and covert crossmodal anchoring* with the function of *perceptual tracking unique distal things*. The crossmodal links should contribute to the keeping track of the uniqueness of the distal target by means of providing a continuous tracking across modalities.

Given that demonstrative identification *of one single object* to keep track of *as long as possible* of the spatio-temporal path of the target object. For spatio-temporal tracking is a reliable method to *secure* the cognitive contact with a single and unique object. Keeping track of this path helps avoid conflating distinct targets.

Now, given that the most frequent interactions with ordinary physical objects are given in distinct pieces of modal information (we look at what we have just heard, we grasp what we have just seen etc.), the use of crossmodal attention seems necessary to recover the *continuity* of the *spatiotemporal path* of each single target.

For example, the event of the fall and impact of the book on the floor occurs with an abrupt sound, visual saccades and vision of the book on the floor: crossmodal attention is needed to link the perceived auditory direction of the impact sound with the visibility of the book at its new place on the floor. The argument for crossmodal attention is that *crossmodal attention seems to rest on routines that improve the cross-modal localization of distal objects/events (within an egocentric frame of reference)*, and therefore facilitate its perceptual track keeping. If this kind of analysis is correct, one should discover many crossmodal links that help in the perceptual tracking of the target object.

5 Epistemic attention as component of multimodal identification

Since the crossmodal anchoring of capture is likely to be performed before any explicit conceptual operation it serves to distinguish this nonconceptual way of tracking an *x* from the conceptual ways of identifying and reasoning about this particular *x*. Part B of proposition 1 refers to the latter.

---

[17] It has been shown in particular that visual lip-reading improves our auditory capacity to recover linguistic information from the acoustic medium, and consequently makes it easier to grasp the content of the perceived speech (Calvert, Brammer, & Iversen, 1998; Driver, 1996; Driver & Spence, 1994; Reisberg, 1978). In other words, it is useful in a context of superimposed speech to pick out not only the relevant speech-sounds from among those entering the ears, but also the matching lip-movements from among the visual stimuli impinging the retina. A speaker is therefore a typical target object for crossmodal attention; and crossmodal attention seems to get multimodal access to simultaneously audible and visible objects like speakers.

A strategy that can lead to the justification of the claim B is to contrast the case of the epistemic use of *one single 'pure' sensory modality* with the case of *crossmodal coupling of information*. The output should lead to the effect that, as compared to any pure modality case, the use of crossmodal coupling is more likely to explain how one identifies the target on the basis of the knowledge of its spatiotemporal path.

I will develop here the example of auditory attention, and contrast the case of '*pure audition*'[18] with the case of the coupling of auditory attention and visual attention.

According to a theory based on attentional procedures, once auditory attention has selected a particular cue related to a distal sound-emitting object, the mind subsequently performs perceptual cycles[19] of procedures (or a hierarchy of procedures for interrogating the distal world) that convert the accessed cue into skills and knowledge. We can use the notion of *epistemic queries* to refer to one particular stage of this attentional cycle, the act of verifying observational propositions. This mental act involves control procedures that enable and determine the search of information for evaluating a concept or a proposition (or cause and justify this search as Campbell (2002) suggests). It depends on the background of particular goals, actions and expectations (but I shall not analyze this background here). Although a number of queries may be non-conceptual routines[20], I assume that at least one set of queries have conceptual content, since they relate to the conceptual identification of the main target source. I assume that one can describe the epistemic queries belonging to the latter conceptual set under a logical or a linguistic form. Described in that way, auditory epistemic queries may respect a form such as:

> Check (or verify, analyze) now with auditory routine *r* whether this tracked individual *x* located here is currently an *F*; or,
>
> Verify [*with particular auditory routines*] that this *x* here is *F*.

The latter form contains a 'that-clause', which makes explicit the identification queries related to what Dretske (1995: 331-35), among others, calls 'fact perception' (perceiving that *x* instantiates the concept *F*), 'meaningful perception' or 'perceiving-*as*'. For instance, resolving the query 'verify that this sound is an ambulance sound' requires, trivially,

---

[18] The phrase *pure audition* refers to audition conceived in isolation from other senses. It parallels the use of the phrase *pure vision* by Churchland, Ramachandran, & Sejnowski (1994).

[19] There is an "attentional cycle" for the emission and resolving of epistemic queries [ref on "perceptual cycle or feedback"] – which is the cycle of the capacity of perceptual scrutiny. It also relates closely with perceptual inferences about the target object.

[20] Cf. for instance Evans (1982) and Peacocke (1992) about nonconceptual routines.

possessing the concept of an ambulance (cf. table 1 in the appendix for other examples of auditory epistemic queries).

Prima facie 'pure audition' routines that can solve epistemic identification queries may contribute to the 'identification' of the form 'this sound/source *x* is *F*', primarily based on information picked up by audition. Once attention has selected a particular sound source, distinct auditory routines may serve to search and recognize relevant properties of the resonating object or the acoustic event. Auditory attention can pick up information about various *properties* or features of the layout of physical objects (e.g., Handel, 1995). These include spatial characteristics of the sound source – e.g., direction, distance, and perhaps gross geometric properties (Lakatos, McAdams, & Caussé, 1997) –, recognition of material (stuff) and mechanical properties of the source, recognition of events between sound sources such as vibrations of solids (such as scraping or rolling), motions of gases (such as exploding balloons or wind), and impacts involving liquids (such as splashing or pouring). Hence, it can be concluded that auditory attention performs property recognition, that is, identification in the weaker sense (cf. section 2).

Nonetheless, the problem remains whether the auditory recognition of *x* being *F* would amount to the genuine demonstrative identification of the particular *unique* object *x* – i.e. the stronger sense of 'identification' (see above). With respect to 'pure audition' a *deflationist account of auditory identification* would claim that auditory perception in isolation cannot identify an *individual* in the strongest sense – akin to Evans' theory (possession of the ability to distinguish the target object from all other objects of the same sort) – since pure audition cannot genuinely keep track of the target object. The following arguments support this deflationist account.

The most important argument is related to (derives its strength from) auditory localization and spatio-temporal tracking. It is a fact that audition furnishes information about source directions in ideal condition (anechoic chamber). Nonetheless, perceiving directions within an egocentric frame of reference cannot be equated with perceiving source *locations* in an objective (or allocentric) frame of reference, for directional information lacks *distance* information and hence cannot locate the target accurately on a particular allocentric map-like structure. In addition, estimation of the source location can be rendered difficult in environmental contexts by echoes and reverberations (Wightman & Jenison, 1995). As a result, if demonstrative identification depends on the *continuous spatiotemporal tracking of the source location* (as Evans suggests), this spatial tracking seems doubtful with 'pure audition'.

The proponent of this view will probably insist on asymmetries between vision and audition with respect to the spatial tracking: vision gets access to located boundaries and surfaces (Kubovy & Van Valkenburg, 2001); audition bears essentially on *dynamical* information about sources but cannot furnish information about the location of boundaries and surfaces of objects, and does not seem to get continuous spatial tracking of successive locations of the target (e.g., the object becomes "absent" to the auditory sense as soon as its ceases to emit sounds). As a result, according to the deflationist view, auditory cognition would be restricted to 'general thoughts' about objects – that is, identification in the weakest sense, which cannot resolve the problem of securing the reference to one *unique* object.

However, it is still possible to dismiss the deflationist account if auditory 'identification' is related to multimodal perception and crossmodal attention. The general reason continues to be that crossmodal attention allows linking the perceptual tracking within distinct sensory modalities, and operates also for epistemic identification. For instance, in the example of the fall of the book on the parquet floor (and exogenous attention) audition and vision can be conceived of as complementary contributions for *locating* and *keeping track* of the (fallen) object that has undergone the event responsible for capture. Auditory capture 'pulls' attention in the direction of the event, but only vision can inform accurately about the *new location* of the object and allows performing identification queries about that individual object. But without audition, most people perhaps would not have been informed about the fall of the object on the floor, especially if the object was not within their field of vision at the time of the occurrence of the event.

6 Conclusion

In summary, the crossmodal view suggests a new framework for the study of demonstrative identification. The study of crossmodal procedures of attending is likely to help in understanding how the mind connects to distal objects, since a number of identification procedures are based on crossmodal skills. This view advocates two theses. First, one routinely uses the coordinating ability of crossmodal attention to retrieve the *continuity* and *uniqueness* of the spatiotemporal path of the target of demonstrative identification. Secondly, crossmodal attention can be of two kinds. First, it may rely on nonconceptual skills that provide the anchoring on the target (as in the example of crossmodal capture). Second, crossmodal attention may be based on epistemic identification routines based on crossmodal tracking. It seems likely that demonstrative identification, in the strongest sense (that relies on

a grip on the spatiotemporal path of the target element for securing uniqueness contact), is better explained as a crossmodal phenomenon.

7 Appendix: examples of auditory epistemic queries

Type of target sources and events		*Examples of auditory epistemic queries in 'pure audition'*
Objects or sources	**Events related to the sources**	**Description of one epistemic query related to the source**
Human individual Speaker	Speech production Vocal emissions	Check whether this voice [*while tracking an invisible individual who is speaking in another room*] is that of *x* or *y* [where *x* and *y* stand for two distinct proper names].
Human individual Speaker	Speech production Vocal emissions	Determine whether these vocal sequences [*previously heard on the right in the dark*] come from the same individual as those vocal sequences [*currently heard on the left*].
Human individual	Motion in internal organs	Verify [*while performing the auscultation of x*] whether the rhythm of *x*'s heartbeat (or breathing) is regular
Human musician Artifact, musical instrument	Resonances of one instrument due to the actions of the musician Playing of a melodic sequence	Check whether this sound [*while listening to the playing of one musical instrument*] comes from a trumpet, a trombone, or another kind of musical instrument.
Human driver Artifact, vehicle with a siren	Siren activity Moving of the vehicle	Determine [*while listening to a vehicle siren*] whether this sound originates from an ambulance, a police car, a fire-fighter truck or any other kind of vehicle.
Ephemeral rain droplets	Impacts of a large number of droplets on the ground	Check whether these [*attention to the process of surrounding droplet impacts*] are light or heavy drops of rain? Is this [*attentional selection toward the surrounding shower*] a light or heavy rainfall?

8 References

Baldwin, D. A. (1993). Infant contributions to the achievement of joint reference. In P. Bloom (Ed.), *Language Acquisition, Core Readings*. Cambridge, MA: MIT Press.

Bregman, A. (1990). *Auditory Scene Analysis: the Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Burks, A. W. (1949). Icon, index and symbol. *Philosophy and Phenomenological Research, IX*, 685.

Butterworth, G., & Grover, L. (1990). Joint visual attention, manual pointing, and preverbal communication in human infancy. In M. Jeannerod (Ed.), *Attention and Performance XIII: Motor Representation and Control* (pp. 605-624). Hillsdale, NJ: Lawrence Erlbaum Associates.

Calvert, G. A., Brammer, M. J., & Iversen, S. D. (1998). Crossmodal identification. *Trends in Cognitive Sciences, 2*(7), 247-253.

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

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.

Dretske, F. I. (1979). Simple Seeing. In D. F. Gutafson & B. L. Tapscott (Eds.), *Body, Mind and Method* (pp. 1-15): Kluwer Academic Publishers.

Dretske, F. I. (1995). Meaningful perception. In S. M. Kosslyn & D. N. Osherson (Eds.), *An Invitation to Cognitive Science: Visual Cognition, Second Edition* (pp. 331-352). Cambridge, MA: MIT Press.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature, 381*, 66-68.

Driver, J., & Baylis, G. C. (1998). Attention and visual object segmentation. In R. Parasuraman (Ed.), *The Attentive Brain* (pp. 299-325). Cambridge, MA: MIT Press.

Driver, J., & Spence, C. (1994). Spatial synergies between auditory and visual attention. In C. Ulmità & M. Moscovitch (Eds.), *Attention and Performance XV: Conscious and Nonconscious Processing* (pp. 311-331). Cambridge, MA: MIT Press.

Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences, 2*(7), 254-262.

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

Evans, G. (1985). Molyneux's Question. In A. Phillips (Ed.), *Collected Papers*. Oxford: Oxford University Press.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. London: George Allen and Unwin.

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

Hatfield, G. (1998). Attention in early scientific psychology. In R. D. Wright (Ed.), *Visual attention* (pp. 3-25). Oxford: Oxford University Press.

James, W. (1890a). Attention, *The Principles of Psychology*. New York: Dover Publications.

James, W. (1890b). *The Principles of Psychology*. New York: Dover Publications.

Jones, M. R. (2001). Temporal expectancies, capture, and timing in auditory sequences. In C. L. Folk & B. S. Gibson (Eds.), *Attraction, Distraction and Action: Multiples Perspectives on Attentional Capture* (pp. 191-229). Amsterdam: Elsevier.

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.

Kaplan, D. (1989a). Afterthoughts. In J. Almog & J. Perry & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 556-614). Oxford: Oxford University Press.

Kaplan, D. (1989b). Demonstratives. In J. Almog & J. Perry & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481-563). Oxford: Oxford University Press.

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition, 80*, 97-126.

Lakatos, S., McAdams, S., & Caussé, R. (1997). The representation of auditory source characteristics: Simple geometric form. *Perception & Psychophysics, 59*(8), 1180-1190.

Logan, G. D., & Compton, B. J. (1998). Attention and automaticity. In R. D. Wright (Ed.), *Visual attention* (pp. 108-131). Oxford: Oxford University Press.

Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.

Malle, B. F., Moses, L. J., & Baldwin, D. A. (Eds.). (2001). *Intentions and Intentionality*. Cambridge, MA: MIT Press.

McAdams, S., & Bigand, E. (1993). *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford: Oxford University Press.

Metzger, W. (1953). *Gesetze des Sehens*. Frankfurt am Main: Waldermar Kramer.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.

Peacocke, C. (1992). *A Study of Concepts*. Cambridge, MA: MIT Press.

Peirce, C. S. (1931-35). *Collected Papers of Charles Sanders Peirce, Vols. I-VI*. Cambridge, MA: Harvard University Press.

Perry, J. (1977). Frege on demonstratives. *Philosophical Review, 86*(4), 474-497.

Perry, J. (1979). The problem of the essential indexical. *Noûs, 13*, 3-21.

Plomp, R. (2002). *The Intelligent Ear: On the Nature of Sound Perception*. Mahwah, NJ: Lawrence Erlbaum Associates.

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.

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

Reisberg, D. (1978). Looking where you listen: visual cues and auditory attention. *Acta Psychologica, 42*(331-341).

Russell, B. (1911). Knowledge by acquaintance and knowledge by description. *Proceedings of the Aristotelian Society, 11*, 108-128.

Russell, B. (1956). *Logic and Knowledge, Essays 1901-1950 (ed. by R. C. Marsh)*. London: George Allen & Unwin.

Russell, B. (1984). *Theory of Knowledge, The 1913 Manuscript*. London: Georges Allen & Unwin.

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

Shiffrin, R. M. (1997). Attention, automatism, and consciousness. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific Approaches to Consciousness* (pp. 49-64). Hillsdale, NJ: Erlbaum.

Spence, C. (2001). Crossmodal attentional capture: A controversy resolved? In C. L. Folk & B. S. Gibson (Eds.), *Attraction, Distraction and Action: Multiples Perspectives on Attentional Capture* (pp. 231-262). Amsterdam: Elsevier.

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

Titchener, E. B. (1899). *An Outline of Psychology*. New York: The Macmillan Company.

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: Lawrence Erlbaum Associates.

Treisman, A. (1969). Strategies and models of selective attention. *Psychological Review, 76*(3), 282-299.

Ullman, S. (1984). Visual routines. *Cognition, 18*, 97-159.

Wightman, F. L., & Jenison, R. (1995). Auditory spatial layout. In W. Epstein & S. Rogers (Eds.), *Handbook of Perception and Cognition, Vol. 5: Perception of Space and Motion*. San Diego, CA: Academic Press.

Woodruff Smith, D. (1989). *The Circle of Acquaintance: Perception, Consciousness, and Empathy*. Dordrecht, Boston, London: Kluwer Academic Publishers.

Wundt, W. M. (1897). *Outlines of Psychology* (C. H. Judd, Trans.). Leipzig: Wilhelm Engelmann.

Yantis, S. (1998). Objects, attention and perceptual experience. In R. D. Wright (Ed.), *Visual Attention* (pp. 187-214). New York, Oxford: Oxford University Press.