On the very idea of a frame of reference

Jérôme Dokic & Elisabeth Pacherie
Institut Jean-Nicod (CNRS-EHESS-ENS), Paris

0. Introduction

It is widely assumed, both in philosophy and in the cognitive sciences, that perception essentially involves a relative or egocentric frame of reference. In his discussion of a variant of Molyneux’s question concerning the relationship between the frames of reference used in particular languages and the frames of reference involved in non-linguistic spatial representations, Levinson has explicitly challenged this assumption. Instead, he argues in favour of the 'neo-Whorfian' hypothesis that the frame of reference dominant in a given language infiltrates spatial representations in non-linguistic, and in particular perceptual, modalities.

Our aim in this paper is to assess Levinson's neo-Whorfian hypothesis at the philosophical level and to explore the further possibility that perception may not just use frames of reference other than the relative one but may also, in some cases at least, be perspective-free, in the sense that it does not involve any frame of reference (whether relative, intrinsic or absolute).

We shall proceed as follows. First, we shall introduce Levinson's variant of Molyneux's question (section 1) and his useful taxonomy of frames of reference (section 2). A logical reconstruction of Levinson's argument will then be given (section 3). The rest of the paper will focus on perceptual and linguistic representations. We shall first discuss four important differences between them, which arguably are compatible with Levinson's argument (section 5). Then we shall ask whether, contrary to what Levinson claims, perception is necessarily perspective-bound (section 6). In the next section, the notion of implicit frame of reference will be introduced, and we shall examine its relevance to the part of Levinson's argument that concerns the compatibility of frames of reference across modalities (section 8).

1. Generalizations of Molyneux's Question

In 1688, William Molyneux wrote a letter to John Locke posing the following celebrated question, which Locke reproduced in his Essay Concerning Human Understanding (1690: II, ix, 8):

---

1 This paper is based on earlier joint work on frames of reference; see Dokic & Pacherie (1999). We thank two anonymous referees for many helpful comments.
2 We adopt Levelt's terminology (1996) and call 'perspective-free' the representations that do not employ frames of reference in Levinson's sense and 'perspective-bound' those that do. Of course there might be other senses in which a 'perspective-free' spatial representation can be perspectival.
Suppose a man born blind, and now adult, and taught by his touch to distinguish between a cube, a sphere of the same metal, and nighly of the same bigness, so as to tell, when he felt one and the other, which is the cube, which is the sphere. Suppose then the cube and sphere placed on a table, and the blind man to be made to see: Quaere, whether by his sight, before he touched them, he could now distinguish, and tell, which is the globe, which the cube.

Nowadays, this question is emblematic of the issue of whether spatial representations are modality-specific or not. In other words, are spatial representations specific to a modality such as vision or touch, or are there supramodal or amodal ways of representing space?

In its original formulation, Molyneux's Question was about representations of shape, but it can be generalized to representations of locations and directions. For instance, we might devise the following variant of Molyneux's Question, which deals with the auditory modality:

Suppose a man born blind, and now adult, and taught to locate in space two auditory sources, say a sphere and a cube placed on a table, one emitting a low note and the other a high note. Suppose then the blind man to be made to see: Quaere, whether by his sight and without hearing anything, he could now tell which objects the notes came from.

In his well-known essay on Molyneux's Question, Evans (1985) in effect tried to reduce the original Question to variants of this kind. Now there is another direction in which Molyneux's Question can be generalized. The questions formulated so far are about spatial representations in sensory modalities. However, non-sensory modalities (or representational systems) also involve spatial representations. In other words, there are conceptual and linguistic representations of space. So we might think of other variants of Molyneux's Question which deal with linguistic representations of locations, such as the following:

Suppose a man born blind, and now adult, is told about the location of a cube and a sphere in front of him. Suppose that the blind man is made to see: Quaere, whether by his sight and without any further testimony, he could now tell where is the cube and where the sphere.

One motivation for treating perception, conception and language as modalities is that there seems to be a unifying concept that can be invoked to describe our representations of locations at all three levels (Levinson 2003: 56-7). This is the concept of a frame of reference. As Pinker (1997: 262) puts it, "reference frames are inextricable from the very idea of location". Indeed, frames of reference are regularly invoked to characterize the structure of spatial maps and the modes of presentation of locations in various modalities.

As far as language is concerned, the expression of spatial relations involves frames of reference which can vary from one language to the other. The issue we would like to discuss

---

3 Levinson's own characterization of modalities is as follows (1996: 152-153): "What we should mean by 'modality' here is an important question. In what follows, I shall assume that corresponding to (some of) the different senses, and more generally to input/output systems, there are specialized 'central' representational systems, for example, an imaginistic system related to vision, a propositional system related to language, a kinaesthetic system related to gesture, and so on."

4 For instance, see the section on frames of reference in Eilan et al. (1993), and especially the introduction by Bill Brewer and Julian Pears. See also Campbell (1994) and Pinker (1997).
concerns the relationship between the frames of reference involved in particular languages and the frames of reference involved in non-linguistic spatial representations, especially in perception. This issue has been addressed in Levinson's work on frames of reference (see Brown & Levinson 1993, 1994; Levinson 1996; Brown & Levinson 2000; Levinson 2001; Levinson 2003). In particular, Levinson gives Molyneux's Question a new twist by linking it with the following questions:

**Levinson's Questions** (1996: 153):
1. "Do the different representational systems natively and necessarily employ certain frames of reference?"
2. "If so, can representations in one frame of reference be translated (converted) into another frame of reference?"

Levinson's own answers to these questions are based on a 'neo-Whorfian' hypothesis. The original Whorfian hypothesis (see Whorf 1956) was that our perception and conception of the world are to a large extent determined by the semantic and grammatical organization of our language. On Levinson's weaker, neo-Whorfian hypothesis, the various modalities can in principle operate in their own, proprietary frames of reference, but Whorfian effects arise because of the need for them to coordinate and exchange information. In Levinson's own words (1996: 157), first, "the frame of reference dominant in a given language 'infiltrates' other modalities, presumably to ensure that speakers can talk about what they see, feel, and so on" and, as a consequence, "other modalities have the capacity to adopt, or adapt on, other frames of reference, suggesting a yes answer to Mr. Molyneux."

### 2. Three kinds of frames of reference

Since its modern introduction by Gestalt psychology, the notion of frames of reference has been widely used in various disciplines, such as linguistics, psychology, philosophy and the neurosciences. One of Levinson's important achievements is his systematic typological work in surveying and simplifying the various distinctions to be found in the relevant literature. In this section, we shall describe what Levinson sees as the three main kinds of frames of reference, namely intrinsic, absolute and relative frames of reference.

[Figure 1: intrinsic, absolute, and relative frames of reference; from Figure 4.9 of Levinson 1996: 139]

Basically, any frame of reference involves the selection of reference objects (the referent and the relatum) and determines the way the spatial relation between them is represented.

The simpler frame of reference, and perhaps the most widespread in natural languages, is the *intrinsic frame of reference*. It allows for the representation of binary relations between referent and relatum. The origin is fixed on the relatum, and the coordinate system is determined by intrinsic properties of the relatum. These properties exploit the relatum's asymmetries and functions, often in a culturally specific way.

---

5 See for instance Whorf (1956: 213): "The categories and types that we isolate from the world of phenomena we do not find because they stare every observer in the face; on the contrary, the world is presented as a kaleidoscopic flux of impressions which has to be organized by our minds – and this means largely by the linguistic systems in our minds."
The absolute frame of reference also allows for the representation of binary relations between referent and relatum (Levinson 1996: 145). Similarly, the origin is fixed on the relatum, but the coordinate system is determined by fixed bearings in the environment, such as the cardinal directions.6

Finally, the relative frame of reference allows for the representation of ternary relations among referent, relatum and point of view. The origin of the primary coordinate system is the point of view, and the coordinate system "seems generally to be based on the planes through the human body, giving us a up/down, back/front and left/right set of half lines" (Levinson 1996: 142).7

It is important to note that the use of each frame of reference involves different cognitive abilities and resources. First, use of an intrinsic frame of reference requires the ability to identify and analyze the relevant relatum – its orientation, functional and dynamical properties (Levelt 1996: 87). Second, use of an absolute frame of reference requires the ability to keep track of the relevant fixed bearings over time (dead reckoning might be necessary). Third, use of a relative frame of reference requires in particular the ability to keep track of one's left and right, which goes hand in hand with the ability to recognize enantiomorphs across perceptual contexts (e.g. to recognize a left hand from one scene to another).

As it is now well-known, particular languages do not favour the same kinds of frames of reference. Levinson points out that some languages predominantly use only one kind of frame of reference, which can be the absolute one or the intrinsic one. It appears that if a given language uses a relative frame of reference, it also uses an intrinsic one. Other languages use two kinds of frames of reference, which can be either the relative and the intrinsic ones, or the absolute and the intrinsic ones. Finally, some languages, including English, Dutch and French, use all three kinds of frames of reference.

3. The case of Tzeltal

Levinson and his collaborators conducted a series of experiments in order to understand the relationship between frames of reference used in linguistic and non-linguistic representations. More precisely, these experiments were designed to test the influence of linguistic frames of reference on the coding of spatial relations in non-linguistic tasks.

One target language of Levinson's experiments is Tzeltal, a Mayan language spoken in Tenejapa (Chiapas, Mexico). The peculiarity of Tzeltal is that it does not seem to involve a relative frame of reference; it has no words for 'left' and 'right'. Tzeltal employs an absolute frame of reference, which is used for coding the spatial relations between objects separate in space, and an intrinsic frame of reference, which is used only for coding the spatial relations between parts of a single object or between contiguous objects.

---

6 Isn't the relation of being north of really ternary, since the environment seems to play a role along with the referent and the relatum? No, because the reference to the environment is built into the lexical meaning of "north" and hence does not act as a separate variable.

7 There can be a secondary coordinate system (often involving the "front"/"back" terms) whose origin is the relatum. Moreover, the point of view itself can be fixed or mobile.
Tzeltal's absolute frame of reference is derived from a topographical feature; Tenejapa is a large mountainous tract which tends to fall in altitude toward the north-northwest. In particular, the term we might translate as 'downhill' designates approximately the northern direction, while the term we might translate as 'uphill' designates approximately the southern direction. There is only one word, which we might translate as 'across', designating indifferently the eastern and the western directions. Note that this frame of reference is still used when Tenejapans are transported outside their territory:

[Figure 2: topographical representation; from Figure 4.1 of Levinson 1996: 112]

In his experiments, Levinson compared the performance of speakers of Tzeltal with that of speakers of Dutch in tasks of recognition and spatial reasoning.

In a typical experiment, the subject sees an arrow on a table. The arrow points to her right, or objectively to the north. The subject is then rotated 180 degrees to face another table. There are two arrows on the second table, one pointing to her right, the other to her left. The subject is asked to identify 'the arrow like the one he saw before' (1996: 114).

[Figure 3: design of some experiments; from Figure 4.2. of Levinson 1996: 113]

The results of these experiments are that Dutch subjects predominantly use a relative frame of reference (i.e. they tend to preserve the orientation of the arrow relative to themselves), whereas Tenejapans predominantly use an absolute frame of reference (i.e. they tend to preserve the orientation of the arrow relative to their environment). In another study, Levinson and Brown (1994) have shown that Tenejapans make no essential use of the notions of left and right in daily life. As a consequence, they are relatively insensitive to left/right inversions of enantiomorphs.\(^8\)

4. A reconstruction of the neo-Whorfian argument

In this section, we would like to reconstruct the main logical steps of Levinson's neo-Whorfian argument (as given for instance in section 4.4 of Levinson 1996 and section 2.4 of Levinson 2003) in order to assess it at the conceptual level. We thus aim at a better understanding of the notion of a spatial frame of reference and of the nature of cross-modal transfer of spatial information.

**Levinson's neo-Whorfian argument**

1. "Any and every spatial representation, perceptual or conceptual, must involve a frame of reference" (Levinson 2003: 56).

---

\(^8\) Levinson's experiments have been criticized, for instance by Li & Gleitman (2002). These authors have argued that speakers of languages using the three kinds of frames of reference can choose any of them depending on the context. For instance, if there are visible landmarks around, some native speakers of English have a preference for an absolute frame of reference, just as speakers of Tzeltal. In their reply, Levinson et al. (2002) point out that Li & Gleitman do not distinguish between intrinsic and absolute frames of reference, and argue that their experimental subjects actually use intrinsic frames of reference, which is fully compatible with Levinson's claim that subjects will tend to use the frames of reference, in that case intrinsic and relative, which are predominant in their language.
2. There is a cross-modal sharing of information. For instance, we can talk about what we have seen, we can gesture about what we have explored by touch, and so on.

3. Such cross-modal sharing of information would only be possible under one of the following conditions:
   3.a. One can translate spatial information from one frame of reference to another.
   3.b. A single frame of reference operates in the relevant modalities.
   3.c. Some modalities can adopt various frames of reference depending on the context.

4. Against 3.a: There are constraints on translation, so that one cannot freely convert information from one frame of reference to another.

5. Against 3.b: It is not the case that all modalities operate with the same frame of reference.

6. Hence, from 3, 4 & 5, some modalities should be able to adopt various frames of reference depending on the context.

7. Some languages, such as Tzeltal, have only one frame of reference (in the case of Tzeltal, an absolute one).

8. Hence, insofar as one can express linguistically spatial representations acquired from other modalities (2), these modalities should be able to adopt the dominant (perhaps unique) linguistic frame of reference.

The argument is valid, so we have to reject one or more premisses if we do not accept the neo-Whorfian conclusion. In what follows, our targets will be the following:

**Contra premiss 1**: Spatial information carried by non-linguistic representations need not be based on a frame of reference.\(^9\) In other words, there is a sense in which non-linguistic spatial representations can be perspective-free. This style of objection will be developed in sections 6 and 7, focussing on perceptual representations.

**Contra premiss 3**: Cross-modal transfer of information need involve neither a unique frame of reference nor translatable frames. The model of translation is misleading to characterize the transfer of information across modalities. This objection will be developed in section 8.

\(^9\) A different point is that at least some linguistic spatial representations do not involve any frame of reference. Consider for instance descriptions of simple topological relations as in "The fly is in the room". It is not clear in what sense such a spatial representation could be said to employ a relative, intrinsic or absolute frame of reference.
5. Four differences between perceptual and linguistic spatial representations

To begin with, let us mention four potential differences between perceptual and linguistic spatial representations.

1. Linguistic representation of space can be as selective as one wants. In contrast, non-linguistic representation of space typically carries nested information (Dretske 1981).

   In perception, spatial information is typically nested in other perceptual information. Normally, one cannot perceive a fact about an object without perceiving many other facts about it. Perceptual information is dense. This holds for spatial relations. Normally, one does not just perceive the distance between the chair and the bottle; one perceives many other spatial facts involving these objects, such as their spatial relations to other objects or to the background.

   On the other hand, conceptual and linguistic representations carry more abstract information and allow for a selective representation of the scene. For instance, the distance between the chair and the bottle can be specified in language (at least in a coarse way; see below) independently of a linguistic representation of other spatial relations involving these objects.

2. Linguistic representation of space is often coarse. Non-linguistic representation of space is analogue (Peacocke 1989).

   In perception, it seems that any value of the spatial dimensions may enter the fine-grained content of the representation. For instance, within certain psychophysical limits, any distance between two objects can be determinately represented. In contrast, our ordinary linguistic descriptions of a visual scene will often be less determinate.\(^{10}\) Of course, we can use precise numerical values, but not all languages and cultures have complex linguistic numerical systems (Gordon 2004).

3. Any region of a represented scene can be directly accessed through perception, whereas some of them can only be indirectly accessed through language.

   There is an indefinite number of different regions and directions in the perceptual field, but they can all be directly identified. For instance, when we perceive the region of the visual field which we call 'up to the left', the phenomenology of our perceptual experience is such that this region is directly accessed rather than represented as the product of two localisations: 'up' and 'to the left'. We could express this fact by saying that orientation in perception is homogenous; all egocentric regions and directions are perceived in the same way, i.e. without the mediation of a system of coordinates. This is not to deny that there are privileged dimensions in perception (such as the gravitational axis). The claim is rather that these dimensions are not used as axes relative to which regions and directions have to be located. In contrast, all languages have only a finite set of lexical items for representing

\(^{10}\) Levinson is well aware of this point, as the following quotation shows (2003: 15): "Take, for example, the metric precision involved in seeing a cup before me, judging its distance from me, and reaching for it — there is nothing like this metric precision in ordinary language locative descriptions."
locations and directions in space; many locations and directions can only be referred to using complex phrases.\footnote{See Levinson (2003), especially chapter 3.}

4. Often, the origin of a coordinate system cannot be dissociated from the figures to be located. In language, such a dissociation is possible.

For instance, if perception represents things from a point of view, namely the perceiver's, the latter is fixed at any given time and cannot be changed. I cannot perceive the world from another's point of view. In contrast, I can describe the bottle as being to your left, even if I perceive it as being to my right. It might be objected that the presence of alternative imagined perspectives is essential for perception to be as of three-dimensional objects in three-dimensional space.\footnote{An anonymous referee raised this objection. The role of imagination in objective perception is further discussed in Brewer (1999: 197).} But this is fully compatible with our point since the kind of imaginative act required for three-dimensional perception is precisely a simulation of possible perceptual experiences each with their fixed point of view. The significance of this point will appear in section 7.

These differences, between perceptual and linguistic spatial representations, are substantial, but arguably they do not compromise the possibility of translation or conversion between non-linguistic and linguistic frames of reference. They do not contradict the general claim that perception and language can use the same kind of frame of reference but, as we shall see in section 8, they suggest that the problem of cross-modal flow does not arise in exactly the terms in which Levinson sets it. Let us now turn to other, potentially more significant differences between perception and language as far as spatial representation is concerned.

### 6. Perception and frames of reference

In what sense, if any, does perception involve a frame of reference? In this section, we shall try to answer this question with respect to each of the three kinds of frames of reference we have distinguished following Levinson. We shall argue that none of the three kinds of frames of reference is essential to perception. We shall further suggest that one possibility is that perception is in and of itself perspective-free (in the sense that it need not involve any of Levinson's three kinds of frames of reference). The reason it may appear perspectival is that perceptual information is reorganized by means of frames of reference at higher levels of cognitive processing.\footnote{This claim is also made by Levelt (1996), although our arguments for it are different.}

#### 6.1. Intrinsic frames of reference

In How the Mind Works, Pinker speaks of frames of reference as "overlaying the visual field". In particular, he argues that "objects themselves can plot out reference frames" (1997: 266-7).

As an example, let us consider the spatial array depicted at the bottom of figure 4 below. One way of seeing this array is by representing the bottle as being in front of the chair.
It may seem obvious that this example and others of the same kind show that perception is perspective-bound, and in particular that it involves intrinsic frames of reference. Such frames of reference are intrinsic in the sense that they exploit functional or geometrical properties of the perceived objects.

However, it is not clear to us that the notion of an intrinsic frame of reference is needed to explain how we see the relation of the bottle to the chair. There is another way of seeing the array, namely by representing the bottle as being next to the chair. Of course there is a difference between seeing the bottle as being in front of the chair and merely seeing it as being next to the chair. The latter relation (x is next to y) is perspective-free and requires no analysis of the objects in component parts (it is enough that they be seen as two different wholes), whereas the former relation (x is in front of y) is perspective-bound and requires such an analysis. Yet it may be argued that insofar as we take into account the existence of geometrical and/or functional properties that are perceived over and above the spatial relations between elements in the scene, perspective-bound relations can be reduced to perspective-free relations. In the case of the bottle in front of the chair, we perceive the internal structure of the chair, which is spatially asymmetrical, so that we perceive the bottle not merely as being next to the chair, but as having various spatial relations to parts of it (this is of course a consequence of the fact noted in section 5 that perception carries nested information). To say that we perceive the bottle as in front of the chair is tantamount to saying that we perceive it as being near a particular side of the chair, namely its front.

Our claim is not that perception never uses intrinsic frames of reference; rather it is that such frames of reference cannot be the basic way in which spatial information is encoded in perception. Indeed, this can be put as a logical point. In order for geometrical properties of an object, such as asymmetries, to be exploitable for the definition of an intrinsic frame of reference, these properties must already be perceptually encoded independently of an intrinsic frame of reference, on pain of infinite regress. Intrinsic frames of reference effects in perception are thus compatible with there being a basic level of perception which is perspective-free.

6.2. Absolute frames of reference

In his *Philosophical Remarks*, Wittgenstein famously claimed that visual space involves a frame of reference:

We can also say visual space is an oriented space, a space in which there is an above and below and a right and left.

Despite the use of terms like 'above', 'below', 'right' and 'left', who would seem to belong to *relative* frames of reference, Wittgenstein insists that the visual frame of reference is not relative to anything, but is *absolute*:

---

14 Of course, relational predicates such as 'near' or 'next to' are contextual in the sense that the context determines the relevant scale, but, as Levinson himself points out (2003: 65 sq), they are not perspective-bound in his sense, i.e. they do not require the use of an intrinsic, relative or absolute frame of reference.
And *this* above and below, right and left have nothing to do with gravity or right and left hands. It would, e.g., still retain its sense even if we spent our whole lives gazing at the stars through a telescope (1975: §206).

Interestingly, Russell defended a similar view in *Human Knowledge*:

At every moment, what is in the centre of my visual field has a quality that may be called 'centrality'; what is to the right is 'dexter', what to the left 'sinister', what above 'superior', what below 'inferior'. These are *qualities* of the visual datum, not relations (1948: 316).\(^{15}\)

Wittgenstein gives the following argument for the claim that visual space involves an absolute frame of reference:

Could'n't we imagine a visual space in which we would only perceive spatial relations, but not absolute positions? [...] I don't believe we could.

In visual space there is absolute position and hence also absolute motion. Think of the image of two stars in a pitch-black night, in which I can see nothing but these stars and they orbit around one another (1975: §206).

Is this argument cogent? Wittgenstein's point holds at the level of what he calls a 'phenomenological language', which he contrasts with a 'physical language'. Phenomenologically speaking, visual space is exactly how it appears to us; there is no relevant distinction between visual space as it appears to us and visual space as it really is. As a consequence, if we can see two stars orbiting around one another without seeing any relational change in the scene, visual space must be absolute. However, it does not follow that, physically speaking, it is absolute. The stars actually change their spatial relations to at least parts of our body (including parts of the retina). There is more to the scene than what appears to us. Our perceptual systems can be sensitive to relational changes without representing them explicitly. As situated perceivers, we can perceptually represent the motion of an object without representing the necessary changes in its spatial relations to some bodily sides. From the point of view of a physicalist theory of visual space, Wittgenstein's example shows at best that we can perceive motion in visual space independently of a *representation* of relational changes. It does not show that perception involves an absolute frame of reference, at least at the most fundamental level.

It might be objected that Wittgenstein's notion of an absolute frame of reference is different from Levinson's, insofar as it has nothing to do with the environment (for instance with cardinal directions). However, the point we raised against Wittgenstein has some bearing for Levinson's position as well. In both cases, absolute directions are not perceptually given. In Wittgenstein's case, they are pre-representationally individuated in relation with the perceiver's own body. In Levinson's case, in order to use an absolute frame of reference, a perceiver must at least initially anchor the coordinate system to some environmental gradient (mountain slopes, prevailing wind directions, celestial azimuths, etc.) and be capable of constantly keeping track of cardinal directions as she moves in space. An explanation of such a capacity cannot assume that these directions are in turn given in terms of an absolute frame

\(^{15}\) On Russell's theory, see Casullo (1986).
of reference, on pain of circularity.\textsuperscript{16} The perceptual identification of a direction across perceptual contexts presupposes a non-absolute way of identifying them within a given context (for instance, using a demonstrative such as ‘this direction’). In other words, the perceiver should have an independent way of perceiving the relevant directions, and thus absolute frames of reference cannot be the most basic way in which perceptual information is encoded.

So far, we have argued that neither intrinsic nor absolute frames of reference can be the basic formats of spatial representations in perception. Many cognitive scientists and philosophers indeed maintain that the most basic form of encoding of perceptual information uses relative frames of reference. But is that really the case?

\textbf{6.3. Relative frames of reference}

The claim that perception uses a relative frame of reference lies deep in a whole tradition of thought, from Kant to phenomenology and recent analytic philosophy. Here are some quotations from recent work:

Our perceptual field has an orientational structure, a foreground and a background, an up and down. [...] This orientational structure marks our field as essentially that of an embodied agent (Taylor 1978: 154; quoted by Evans 1982: 156).

[Let us] reflect upon how we might specify the spatial information which we imagine the perception to embody. The subject hears the sound as coming from such-and-such a position, but how is the position to be specified? Presumably in egocentric terms (he hears the sound as up, or down, to the right or to the left, in front or behind). These terms specify the position of the sound in relation to the observer's own body; and they derive their meaning in part from their complicated connections with the subject's actions (Evans 1982: 155).

It may seem like a plain phenomenological fact that we always perceive objects as being to the left, to the right, above or below. However, we think that another of Levinson's great merits is to have shown, or at least suggested, that what we consider as a necessary condition of perception is in fact quite contingent. As Brown and Levinson put it:

Kant was wrong to think that the structure of spatial regions framed on the human frame, and in particular the distinctions based on left and right, are in some sense essential human intuitions (1994: 9).

In particular, speakers of Tzeltal do not possess the notions of left and right. Of course, the fact that they lack these notions does not yet show that they do not perceive the world as being left-right oriented. However, once it has been shown that our cognitive scheme is not necessarily sensitive to distinctions based on left and right, the claim that perception must involve a relative frame of reference loses much of its force.

\textsuperscript{16} If a subject had no other means of referring to directions than through an absolute frame of reference, he could not meaningfully wonder whether, or assert that, a certain perceptually presented direction is North. His assertions could only amount to tautologies, such as ‘North is North’, or contradictions, such as ‘West is North’.
What are the arguments for the Kantian claim that perception must involve a relative frame of reference? One argument is that we need such a frame to recognize enantiomorphs across perceptual contexts. However, the case of Tzeltal shows that we can perceive the world while being relatively indifferent to enantiomorphs. Moreover, recognition across perceptual contexts might involve representations that are not strictly perceptual. The claim that at some level we *encode* perceptual scenes using a relative frame of reference does not show that, at the most basic level, perceptual information is itself organized in such a frame. If it were, indifference to enantiomorphs would indeed be difficult to explain.

Another common argument in favour of the claim that perception must involve a relative frame of reference concerns the connection between perception and action. Evans writes that "egocentric terms are the terms in which the content of our spatial experiences would be formulated, and those in which our immediate behavioural plans would be expressed" (1982: 154). However, it is not clear that perception has to use a relative frame of reference in order to *distinguish* directions in the perceptual scene, such as the left/right direction as opposed to the right/left direction. In each perceptual context, the relevant distinction can be drawn in *demonstrative* terms, for instance as the direction that goes from *here* to *there*. In general, acting on a location in the perceptual scene can be direct, in the sense that it does not rely on an explicit identification of the acted upon location as opposed to other locations in the scene.17

In conclusion, although we naturally *specify* the contents of perception in egocentric terms, perception itself need not involve a relative frame of reference. To this extent we agree with Levinson. However, his point is that since the Tenejapans do not have a relative frame of reference in their language, their perception uses an absolute one. Our point is rather different. To say that in some cases perception does not use a relative frame of reference cannot imply that it uses either an intrinsic or an absolute one, since, as we argued in sections 6.1 and 6.2, these latter frames of reference cannot themselves be basic on pain of infinite regress or circularity. Therefore, Levinson cannot conclude that when perception is not organized at the most basic level in terms of a relative frame of reference, it must use either an intrinsic or an absolute one (the latter in the case of the Tenejapans). Instead, one should conclude that perception, at its most basic level, is either perspective-free or relative and that other kinds of frames of references are imposed on the perceptual data at higher levels of cognitive processing.

### 7. Implicit frames of reference

We have argued that perception can be perspective-free, in contradiction to the first premise of Levinson's reconstructed argument in section 4. This might seem strange, given the wide agreement in cognitive science that "directed action requires the brain to encode the positions of objects in multiple allocentric frames and multiple egocentric frames" (Gallistel 2002: 322). Here we focus on egocentric frames allegedly present in perception. Are these frames the same as Levinson's relative frames of reference?

It is in fact necessary to distinguish two notions of frames of reference. Perception need not involve a relative frame of reference in the sense that what it represents would have to be described by means of egocentric terms such as 'left' and 'right'. Perception is not

17 For further development of this idea, see Dokic (2003).
necessarily relative in this sense. However, it can be said to be relative in a second, different sense. When we perceive objects and other spatial entities, we exploit our actual spatial situation relative to them. This relative situation need not be represented as such in perception. All perception is implicitly egocentric; it represents only local objects, regions, and directions, but it need not represent them as local. The implicit egocentricity of perception is well described by Perry in the opening passage of his essay "Thought without representation":

I see a cup of coffee in front of me. I reach out, pick it up, and drink from it. I must then have learned how far the cup was from me, and in what direction, for it is the position of the cup relative to me, and not its absolute position, that determines how I need to move my arm. But how can this be? I am not in the field of vision: no component of my visual experience is a perception of me. How then can this experience provide me with information about how objects are related to me? (1993: 205).

In the next section, we shall try to go some way toward alleviating the perplexity Perry expresses here. In the meanwhile, the point is that even 'selfless' perception exploits at a pre-representational level one's spatial situation relative to the perceived scene. Perception represents things from a point of view which itself is not normally represented in our experience.

We can use the notion of a frame of reference to capture the second sense of the claim that perception is relative. Perception involves a frame of reference centred on a point of view, namely the perceiver's body, and whose axes are defined on the basis of our bodily axes. However, it is important to realise that this notion of a frame of reference is very different from the one used by Levinson. He assumes that distinctions between linguistic frames of reference can be aligned with distinctions between different formats of perceptual representations. He acknowledges that "at the level of perception, origin and coordinate systems presumably come pre-packaged as a whole, but at the level of language, and perhaps more generally at the level of conception, they can vary freely and combine" (2003: 54). But he suggests that in order to realign a linguistic with a perceptual relative frame of reference, it is enough that we set the origin of the linguistic coordinate system so that it coincides with ego. This may indeed be enough to make the two frames of reference extensionally equivalent. Yet, Levinson fails to grasp the full import of the fact that the origin is fixed in the perceptual case, and in particular its intensional significance.

In language, a relative frame of reference allows for the representation of ternary relations among referent, relatum and point of view. In contrast, in perception the origin is necessarily implicit in the sense that it cannot be part of what is represented. Of course there are cases in which one visually perceive parts of one's own body, but in such cases these bodily parts do not function as the origin of the relative frame of reference; rather, they are themselves located in the visual field relative to an implicit origin. The relative frame of reference used in perception is, as John Campbell puts it, intensional, in the sense that "when the subject is identifying places egocentrically [in perception], he cannot be thought of as doing so by first identifying a physical thing, himself, through a body image, and then identifying places by their relation to his body" (1994: 13). In other words, the egocentric localisation of a position in space does not depend on a prior identification of a body and does not presuppose an explicit representation of oneself as a term of a spatial relation to the position. Of course it must be possible to exploit spatial relations between parts of our body
and the world in order to perceive anything, but this is a condition of possibility of spatial representing, not something that is itself spatially represented. If there is a frame of reference here, it is implicit, not part of what is explicitly represented. To use a linguistic analogy, implicit frames of reference pertain to the syntax of perception, and only indirectly to its semantics.\footnote{For further thought on these lines, see Pacherie (2003).}

So we can maintain the claim that perception need not involve any explicit intrinsic, relative or absolute frame of reference and in that sense can be perspective-free. Moreover, the possibility of implicit frames of reference has an interesting bearing on Levinson's claim about the commensurability between different kinds of frames of reference. To this last issue we now turn.

\section*{8. The cross-modal flow of information}

According to Levinson, the cross-modal flow of spatial information presupposes shared frames of reference: "we will not be able to exchange information across internal representation systems that are not based on one and the same frame of reference" (1996: 155). Indeed, a crucial premiss of the neo-Whorfian argument is that one cannot freely convert information from one frame to another. Levinson gives the following summary of the compatibilities and incompatibilities of frames of reference (Levinson 1996: 154):

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Figure 4: slide from Figure 4.11, Levinson (1996: 154).}
\end{figure}

For instance, the representation 'bottle in front of chair', involving an intrinsic frame of reference, cannot be used to ground the representation 'bottle to right of chair', involving a relative frame of reference. In general, the only directions in which conversion is possible are from the two orientation-bound frames (relative and absolute) to the orientation-free one (intrinsic), and then only if the orientation of the ground object is fully specified.

We have two remarks on this premiss of Levinson's argument. The first is fairly obvious, but perhaps worth reminding. When Levinson argues for untranslatability across frames of reference, he explicitly asks us to 'discard other information'. It is only if we do so that we cannot generate, for instance, the relative description 'bottle to right of chair' from the intrinsic description 'bottle in front of chair'. Such limitations occur only when other information is discarded. In contrast, appropriately nested representations are convertible. For instance, one may use a relative or an intrinsic mode of representation and nevertheless encode cardinal directions systematically enough to allow for a translation into an absolute mode of representation.

The point is relevant to the case of Tzeltal. As we saw, the Tenejapans cannot perceive the world exclusively in terms of an absolute frame of reference. Their ability to keep track of absolute directions cannot be explained unless these directions are perceptually presented, at the most basic level, in a non-absolute way. Now, keeping track of absolute directions enables them to conceive and talk about what they perceive, by converting perceptual information into absolute specifications and descriptions. There is no need to suppose that their language influences the way in which representations are coded at the perceptual level. As far as translatability constraints between frames of reference are concerned, perceptual
representations might involve any kind of relative frame of reference, or indeed, at the most basic level, no frame of reference at all.

Our second remark is more general, and concerns Levinson's model of cross-modal sharing of spatial information. Information flows from a source representational state to a target representational state, i.e. from one representing to another. Now Levinson seems to assume that the target representational state can only exploit what is explicitly represented in the source state. Thus, he uses the image of translation to capture what is involved in cross-modal flow of information. However, Levinson's assumption is unwarranted. Even if a source representational state has a content which cannot be translated into a target representation involving a given frame of reference, there can be enough information associated with the state (although not part of its content) to ground a reliable perspective-bound representation.\(^{19}\) In other words, relevant information which is not explicitly represented in the source state can be implicitly nested in, or associated with it.

Suppose for instance that I perceive a bottle next to a chair. We have argued that my perceptual representation might be perspective-free; in particular, I do not need to perceptually represent the bottle and the chair \(as\) bearing different spatial relations to parts of my body. As a consequence, there is no way of translating what is represented in perception ('bottle next to chair') into a relative description (either 'bottle right to chair' or 'bottle left to chair'). It does not follow that no relative description can be grounded on perception. The perceptual state supervenes on various mechanisms that carry information about spatial relations between elements in the scene and bodily sides. Such information is not part of what is perceptually represented, but can be exploited by a cognitive system in order to produce the perspective-bound representation 'bottle right to chair'. A conceptual representation based on perception can be sensitive not only to what is perceptually represented in perception, but also to the \(mode\) or \(manner\) of perceiving the scene.

In a nutshell, the cross-modal flow of information requires \textit{reliable} transitions between representational states. A reliable transition can be sensitive to more than what is explicitly represented in the source state.\(^{20}\) In particular, it can be sensitive to information implicitly nested in such a state or carried by underlying mechanisms. Levinson's model of translation is too crude to characterize the transfer of information from one mode of representation to another. Information can be exchanged even if there is an incompatibility at the level of content or what is explicitly represented.

From these two remarks, we conclude that there is no good argument from translation limitations to neo-Whorfianism, and that Levinson's third premiss (section 4) does not hold.

\(^{19}\) This idea that information implicit at one level can be made explicit at higher levels is at the heart of the theory of representational redescription (Karmiloff-Smith 1992) and of the account of implicit knowledge given by Perner & Dienes (1999) in various cognitive domains.

\(^{20}\) This point also holds of other kinds of transition. For instance, the transition from the belief that it's raining to the belief that one believes that it's raining can be reliable (it is an instance of reflection), and is sensitive both to what is explicitly represented (that it's raining) and to the mode of the representation (the fact that what is represented is believed). For an account of such a transition, see Peacocke (1999).
9. Conclusion

Although we agree with Levinson's negative point that perception need not essentially make use of a relative frame of reference, we do not think he succeeds in establishing his positive neo-Whorfian claim. There are two alternatives to the view that perception essentially involves a relative frame of reference. One is that perception does indeed essentially involve a frame of reference, but that this frame can also be absolute or intrinsic. The other is that perception can be perspective-free, in the sense that it need not involve any explicit frame of reference. Levinson considers only the first alternative and claims that perception tends to adopt the frame of reference dominant in the language of the perceiver. We argued that this cannot be generally the case, given the dependent nature of intrinsic and absolute frames of reference. This clearly favours the second alternative. We also argued that once the existence of implicit, as opposed to explicit, relative frames of reference in perception is taken into account, the transfer of information across modalities can be explained without appealing to Levinson's neo-Whorfian hypothesis.

REFERENCES

**Figure 1**: Intrinsic, absolute, and relative frames of reference; from Figure 4.9 of Levinson 1996: 139.
Figure 2: Topographical representation; from Fig. 4.1 of Levinson 1996: 112.
**Figure 3:** Design of some experiments; from Figure 4.2. of Levinson 1996: 113.
Figure 4: Untranslability across frames of reference; from Figure 4.11 of Levinson 1996: 154.