Logical constructions suggested by vision

Michiel van Lambalgen[1]
Institute for logic, language and computation
University of Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam
michiell@wins.uva.nl

November 26, 2002

[1]This research was supported by the Netherlands Organisation for Scientific Research (NWO) under grant PGS 22-262. I thank Jaap van der Does for numerous helpful conversations.

0.1 Veridicality

The aim of this paper is to inquire into the possibility of a *logic* of (visual) perception, and to relate this logic to the semantics of perception reports. There are two main reasons why this endeavour goes against the trend in current semantics: it is ordinarily held, following Lewis [12], that considerations of psychology and pragmatics should be kept separate from semantics, and furthermore that in any case, due to the special nature of perception, psychological mechanisms governing perception cannot influence the logic of perception reports. This view is clearly stated in the following quotation from Barwise and Perry [1]

> Seeing is clearly a causal relation, but also, and more importantly, an information preserving one. The other attitudes we study are sensitive to how the information or misinformation actually influences the agent's thought and action. By contrast, epistemically neutral reports are concerned only with the nature of the situation about which information can be detected by the agent.

To put it succinctly, since 'seeing' is information preserving, the semantics of perception reports should be formulated in terms of parts of the *objective* world. Barwise captures the resulting logic of perception reports in a number of principles, the most important of which is, for our present purposes, the principle of veridicality:

> For simple naked infinitive sentences φ, if a sees φ, then φ.

Although Barwise admits that nonveridical readings are possible, it is the intended veridical reading which has to be unravelled; 'having done so, we can then back up and see what modifications would be necessary to accomodate nonveridical readings'. This tenet at least of situation semantics is uncontroversial in the linguistics literature. However, if we now turn to the psychological literature, we see a very different picture. Here is a representative quotation from Miller and Johnson-Laird's *Language and Perception*

> 'See' is one of the more complex verbs in English. Its complexity arises partly from the complicated logic of perception [...] No claim about veridical visual perception can be visually verified. The individual himself is the only person to know what it is that he perceives; other persons cannot verify this component of a veridical claim. But the individual himself can have no grounds for moving from an opague and potentially nonveridical report to a transparant and veridical report. It seems that SEE is a concept that has truth conditions that are easy to state yet impossible to execute. Yet it is possible to determine that an empirical claim is false. [...] Perceptual statements have empirical content, but, like scientific conjectures, they may only be falsified. The output of the perceptual system is a hypothesis about the world; there are no facts except in the light of hypotheses. (Miller and Johnson-Laird [14])

Even limited knowledge of the intricacies of visual information processing would suffice to assent to the truth of Miller and Johnson-Laird's observation; yet it goes without saying that Barwise and Perry are competent speakers of English as well. A possible source for this apparent contradiction is that the two theories consider different readings of 'see' to be the most prominent one. It is clear from the above quotation that Miller and Johnson-Laird emphasise the use of 'see' as it occurs in *first* person perception reports. 'I see φ' means that on the basis of my visual impressions I judge φ to be the case. In this case it is for psychology to decide whether the perception report must necessarily be veridical, and present-day psychology sides with Miller and Johnson-Laird here. Barwise, however, appears to emphasise the use of 'see' as it occcurs in *third* person perception reports. When can I judge truthfully that a sees φ? If φ is true and if a saw it. Clearly on this reading veridicality is part of the meaning of 'see'. In this case the perceptual content φ is ascribed to a by an outside observer who faithfully describes the scene that a is looking at. It is not implied that a was aware of, or believed in φ. (This is what is called an 'epistemically neutral' report.) The sentence 'a sees φ' differs in meaning from 'a sees *that* φ', in that the latter sentence does carry the implication that a was also aware of, or believed that φ. Notice that in the case of first person perception reports, the semantic difference between 'I see φ' and 'I see that φ' is much less marked. Here, both statements normally carry the implication of awareness, one states one's beliefs. If proof would be needed here, one could refer to the phenomenon of 'blindsight': patients with this condition are completely blind, yet can distinguish between a large variety of visual stimuli when asked to do so. However, these patients are surprised by their own abilities and they vigorously deny having seen anything. It is true that one can say, 'I must have seen it, although I was not aware of it at the time', (looking at oneself as an outside observer, as it were) but no one would use such a sentence in the present tense.

Barwise's analysis of perception reports rests in part on a theory of Dretske [4], who argued that beliefs are not necessary for seeing: 'if perception is understood as a creature's *experience* of his surroundings, then perception itself is cognitively neutral' (Dretske [4], p. 153). This seems plausible enough, after all even the housefly perceives although it is not noted for its beliefs. The question is rather, whether this theoretically motivated use of 'see', not involving beliefs, plays a prominent role in natural language. The usual argument given is that we must be able to account for the intelligibility of sentences of the following type

> John sees his daughter in the crowd, although he does not recognise her.

One might analyse these examples differently, however, in terms of cancelled implicatures. Compare the following example

> John sees his daughter in the crowd, and calls his wife.

Everybody would interpret this sentence in such a way that John's calling his wife has something to do with recognising his daughter *as* his daughter. This shows that even in this use of 'see' it is implied, albeit nonmonotonically, that awareness of the perceptual content occurs. If this is true, then the first person reading involving beliefs is more prominent than the epistemically neutral reading. It then follows that veridicality is not in general a valid principle for 'see'.

This analysis is largely in agreement with Jackendoff's treatment of 'see' in [10], where he argues that the meaning of 'see' is described by two nonmonotonic rules, at least one of which has to apply: x sees y if

1. x's gaze goes to y
2. y comes to x's visual awareness.

Stereotypical, veridical, seeing is characterised by the satisfaction of both the first and the second, epistemically positive, condition, but one may felicitously use 'see' when only one of the implicatures obtains. Only when both implicatures are cancelled, as in

*Bill saw a ghost, although he didn't notice it at the time.

is the use of 'see' unacceptable.

Not surprisingly, Jackendoff remarks that his analysis was inspired by Miller and Johnson-Laird. In their work one finds [14, pp. 583–618] the germs of a logical description of first person perception reports, together with a rudimentary analysis of logical principles, such as exportation of the existential quantifier. SEE is characterised in terms of a predicate PERCEIVE(x,y), which holds between a person x and a percept y; SEE$(x,A(y))$ is defined as PERCEIVE$(x,y) \wedge T(y)$, where T indicates the satisfaction of perceptual tests correlated with A. The use of T makes SEE irreducibly opaque, but evolution has seen to it that our perceptual tests are mostly adequate, so that in practice perception reports can be taken to be transparent, veridical.

It is our intention here to follow Miller and Johnson-Laird's lead and to provide a psychologically motivated semantics for 'see'. This will be done by paying careful attention to the nature of T, as discussed in modern theories of vision. In particular, we base ourselves on an abstract account of David Marr's theory of vision [13]. Of course, basing one's semantics on an empirical theory brings with it the danger that the empirical theory is wrong. This is a danger only when one looks for definite answers. A more interesting way to describe this situation is that the psychology of vision and the semantics of perception reports may mutually influence each other, so that for instance a psychological theory may be called into doubt because it gives the wrong semantic predictions.

An issue not addressed by Miller and Johnson-Laird is whether the logic of perception *reports* is really completely determined by the logic of perception; or rather, they assume that this is so. Language is also a social phenomenon, however, so there may be other determinants of the meaning of 'see'. Those who are convinced by Wittgenstein's 'private language' argument may even wish to object to our procedure on the grounds that it is *impossible* that 'see' gets its meaning via psychology. We leave this matter open, as we have nothing particularly intelligent to say here.

0.2 David Marr on vision

We start by explaining the psychological motivation underlying the model theory, taken from David Marr's book *Vision*.

One of Marr's fundamental ideas is that vision is in many ways approximate. Filtering takes place at many of the earlier levels of visual processing, for instance in edge detection, leading up to the so called primal sketch which contains information about planar geometric relationships. Stereopsis and related processes provide some depth clues, ultimately resulting in the $2\frac{1}{2}$ sketch, which represents objects from a viewer-centered perspective. So far perception is bottom-up, with computations on the image proceeding autonomously. However, in order to obtain a full 3D picture of the world, in which the coordinate system is object-centered, a top-down process takes over.

Seeing (in the sense of recognising) a 3-D object involves two processes: constructing an image from visual data as indicated above, and matching the image to a *stored* catalogue of 3-D models, where the matching is based on some salient features derived from the image. For reasons of computational efficiency, this catalogue is built hierarchically, or in a *modular* fashion

> Modularity [...] allows the representation to be used more flexibly in response to the needs of the moment. For example, it is easy to construct a 3-D model description of just the arm of a human shape that could later be included in a new 3-D model description of the whole human shape. Conversely, a rough but usable description of the human shape need not include an elaborate arm description. Finally, this form of modular organisation allows one to trade off scope against detail. This simplifies the computational processes that derive and use the representation, because even though a complete 3-D model may be very elaborate, only one 3-D model has to be dealt with at any time, and individual 3-D models have a limited and manageable complexity ([13, p. 307]).

The important point is that we need many representations of, say, the human shape, some coarse, some more detailed; and that we must be able to keep track of where an elaborate description would fit in the rough model ('indexing'). This is well-illustrated by the following picture (figure 1 from [13, p. 306])[1]

An adequate indexing mechanism allows us to capitalise on the interplay between the clues derived from an image and the matching process (cf. [13, p. 321]): after a 3-D model has been selected (guided by the image), it can be used to search for additional clues in the image; in turn, these can be used (when necessary) to match the image to a more detailed 3-D model. However, it may turn out to be impossible to find a more detailed 3-D model *of the kind we expected*. Indeed, like all computationally efficient heuristics, the use of such

[1] Reprinted by permission from D. Marr and H.K. Nishihara, "Representation and recognition of the spatial organisation of three-dimensional shapes", *Proc. Royal Soc. London B 200*, 269-294.

Figure 1: *Refinement of an arm*

approximate models brings with it the possibility of error: what is identified as a real arm with respect to a given approximation may turn out to be something else (e.g., a wooden arm) when 'looking closer', i.e. with respect to a more refined approximation. A theory such as Marr's is well-suited to account for the hypothetical character of perception emphasised in 0.1: due to our finite perceptual resources, one cannot match the processed image to an arbitrarily refined 3D model. At some point we have exhausted the information in the image and from then on it is only a hypothesis that the object is what we judged it to be. For our purposes it is useful to slightly reformulate the latter idea, as follows: in principle we should be able to match the image to ever more refined 3D models of the object. E.g. if we come closer, the resolution of the image will increase, and hence the possibilities for matching; the *expectation* is that these increased possibilities will not alter our judgement. Of course, it is impossible to obtain complete perceptual knowledge of an object, but since there is no *a priori* bound on what is possible, it seems best to grant us infinite possibilities of refinement, for example an infinitely extendable hierarchy of 3D models.[2]

0.3 Inverse systems and inverse limits

We now turn to a formalisation of the above ideas. Consider again Marr's suggestive example of 3D models of a human. Viewed abstractly, what we see here is a series of first order models, composed of objects and relations between them, together with a mapping specifying how an object occurring at one level is decomposed at the next. This situation can be represented abstractly by means of an *inverse system* of first order models. The basic ingredient is the following:

Definition 1 *Let T be a set directed by a partial order $\geq$; i.e., $\geq$ is reflexive, transitive, anti-symmetric, and for $s, t \in T$ there is $r \in T$ such that $r \geq s, t$. An inverse system (indexed by T) is a structure $\langle \mathcal{M}_s, h_{st} \rangle_{s,t \in T}$ with*

1. *for each $s \in T$, $\mathcal{M}_s$ is a model for the signature σ_s;*
2. *for any R in the union of the signatures there is $t \in T$ such that R is in σ_s if $s \geq t$.*
3. *for each $s, t \in T$ with $s \geq t$ there is a mapping $h_{st} : |\mathcal{M}_s| \longrightarrow |\mathcal{M}_t|$ with for each R in $\sigma_t \cap \sigma_s$*

   $$(\rho) \quad \{\langle h_{st}(d_1), \ldots, h_{st}(d_n) \rangle | \langle d_1, \ldots, d_n \rangle \in R_s\} \subseteq R_t,$$

   where $R_s(R_t)$ is the interpretation of R on $\mathcal{M}_s(\mathcal{M}_t)$
4. *h_{rr} is the identity on $\mathcal{M}_s$, and for $s \geq t \geq r$, $h_{sr} = h_{st} \circ h_{tr}$.*

Here $|\mathcal{M}|$ denotes the domain of $\mathcal{M}$. The mappings h_{st} will be called *bonding* mappings. An inverse system is *total* if in addition the bonding mappings are surjective; we then have the usual concept of homomorphism between models.

Condition (ρ) captures the idea that if $s \geq t$, then $\mathcal{M}_s$ is more refined than $\mathcal{M}_t$. It is mainly for conceptual reasons that we allow the signatures of the models to vary, since we may wish to say that a predicate is not yet applicable at a certain stage. For instance, in a rough approximation of 'human being' given by six appropriately positioned cylinders representing torso, arms, legs and head, the predicate 'hand' is not applicable. Yet, once introduced a predicate should remain applicable. We shall comment on the role of surjectivity in section 0.3.1.

If the inverse system stands for something like the hierarchy of 3D models, then each model in the system can be said to approximate reality. If we make the idealisation that we are in principle able to approximate reality arbitrarily closely (even though never completely), then it seems reasonable to assume that reality is the limit of the inverse system, in the following sense

[2]Cf. the remarks on Husserl's philosophy of perception in [17].

Definition 2 *Let $\langle \mathcal{M}_s, h_{st}\rangle_{s,t\in T}$ be an inverse system. Its inverse limit*

$$\mathcal{M} := \lim_{\leftarrow T} \mathcal{M}_t$$

is defined as follows

1. *the domain $|\mathcal{M}|$ consist of the* threads *in the product $\Pi_{t\in T}|\mathcal{M}_t|$; i.e., functions $\xi : T \longrightarrow \bigcup_{t\in T} |\mathcal{M}_t|$ satisfying: $\xi_t \in |\mathcal{M}_t|$, and $h_{st}(\xi_s) = \xi_t$ for $s \geq t$.*

2. *the interpretation of the predicates is given by: for each R there exists $t \in T$ such that for all threads $\xi^1, \dots, \xi^n$*

$$R(\xi^1, \dots, \xi^n) \Longleftrightarrow \forall s \geq t : R_s(\xi^1_s, \dots, \xi^n_s)$$

The inverse limit $\mathcal{M}$ is a submodel of the direct product $\Pi_{t\in T}\mathcal{M}_t$ (Chang and Keisler [3, p. 224]); however, the domain of this submodel might be empty. Under the additional assumption that the $\mathcal{M}_s$ are finite this cannot be so. The proof rests on the fact that the discrete topology on the $\mathcal{M}_s$ is compact and makes the bonding mappings continuous. (Engelking [5, p. 141].)

Theorem 1 *Suppose $\langle \mathcal{M}_s, h_{st}\rangle_{s,t\in T}$ is an inverse system of finite models. Then $|\mathcal{M}|$ is non-empty.*

0.3.1 Inverse systems as cognitive structure

The introduction of inverse systems is only the first step toward a psychologically realistic and empirically adequate semantics of perception reports; a second ingredient will be introduced in the next section. However, let us step back for a moment and consider the cognitive significance of the formal notions introduced so far. The proposed model contains a number of assumptions which may be thought questionable, so we shall try to be explicit about them.

The first assumption concerns the notion of object. An object in a model belonging to the inverse system is in some sense a *type* of a real, individual object, the *token*; here, type is taken with respect to structural properties such as shape. The hierarchy of 3D models provides a collection of such types, with appropriate indexing relations. In this case refinement often consists in adding more parts, as when an arm is decomposed in upper arm, forearm and hand.[3]

Now consider figure 2 on the following page, also taken from Marr [13, p. 319].

Here, refinement rather consists in adding dimensions. For example, in the case of 'biped' the dimensions of limbs and torso are left unspecified; 'human' and 'ape' are then differentiated at a next stage by a rough indication of length and width in both cases. The associated homomorphism can be viewed as the

[3]See van der Does and van Lambalgen [17] for the definition of a *Marr-model*, the formal correlate of this part-of structure.

Figure 2: *Stick figures*

projection of a vector space onto one of smaller dimension. A number of possibilities of this type are discussed in section 0.6.1 in the context of spatial prepositions. One can think of other possibilities as well. For instance, s may be more refined than t (i.e. $s > t$) because the signature of the model $\mathcal{M}_s$ is larger than that of $\mathcal{M}_t$; in this case h_{st} can even be the identity. Another interesting possibility arises when we take into account the fact that the dimensions characterising a shape (e.g. length of limb, angle between limb and torso) may be known with variable precision, so that the precision itself must be an element of the representation (cf. Marr [13, p. 309]). One way of doing this is to cut up a dimension into finitely many pieces, thus defining an equivalence relation on vectors. If s corresponds to a refinement of the partition at t, then the mapping h_{st} can again be taken to be a homomorphism. Taking all these possibilities together it seems not unreasonable to assume that a token can be viewed as a converging sequence of types. For example, assuming for a moment that an individual human is completely characterised in terms of a shape description (which may include detail at the level of fingerprints), then one can view this individual as also being characterised by a converging sequence of approximate shape descriptions, each of which is applicable to more than one token. (Note that, conversely, a type can be viewed as an equivalence class of tokens, namely the set of those tokens which are indistinguishable from a given token at a certain level of approximation.)

However, can *all* these approximate shape descriptions, or approximate types generally, be said to be part of cognitive structure? Perhaps not in the sense that they must belong to a stored catalogue. But we are able to construct these representations when the need arises, so potentially they *are* part of cognitive structure. The inverse system then represents cognitive structure in this potential sense. The fact that we identified reality with the inverse limit embodies the assumption that, potentially, reality is arbitrarily closely approximable. This assumption lies at the heart of our treatment of veridicality, as will be seen later. [4]

We are now in a position to clarify the role of the surjectivity of the bonding mappings. It is a theorem that if all bonding mappings are surjective, then so are the projections from the inverse limit to the approximating models. It follows that for any approximating model $\mathcal{M}_s$, all objects in $\mathcal{M}_s$ are approximations of a real object. This means that hallucinations are not accounted for. In order to do so, one must allow that it may not be possible to further refine a certain type, i.e. one must drop surjectivity.

The second main assumption concerns the behaviour of the predicates. The reader will have noted that in the way we have set things up, predicates always overgeneralise, in the sense that for each stage s, the set of threads ξ such that $\mathcal{M}_s \models A(\xi_s)$ always includes the set of threads ξ which satisfy A on the inverse limit. For example, in Marr's picture of the biped bifurcating in human and ape, we may wish to interpret the predicates 'human' and 'ape' already at the level of the biped-type. The justification for this move will be discussed in fuller detail in section 0.6.2. In brief, it is a possible model theoretic correlate of the *cascade* theory of the relation between structural and semantic representations, which says that during processing of structural information, simultaneously semantic representations are activated; in particular, that upon seeing an object ('ape'), the semantic representations of structurally similar objects ('human') become activated as well. If object recognition is conceived of along the lines of Marr, this translates into the following picture: upon (vaguely) seeing an ape, the structural representation corresponding to 'biped' is activated, and this in turn activates the representations of its possible continuations ('human', 'ape', 'straw man', ...) and their semantic representations. We formalise this by saying that 'biped' activates the semantic representations of 'human', 'ape', ... directly. This construction explains why we may mistakenly identify, on the basis of incomplete dimensional information, a biped-shape as a human. Some evidence for the cascade theory is presented in section 0.6.2.

[4] Without it, what is perceivable would only be a homomorphic image of reality.

Before we close this section, we add a few words on the role of infinity in these constructions. Interesting examples of inverse systems and limits arise only when the index set is infinite; the reader might object that this makes the enterprise devoid of cognitive meaning. For instance, wouldn't it imply that we need an infinite amount of information about an object before we can recognise it? Yes and no. Yes, because the set up has been chosen in such a way that at any 'finite' stage a perceptual judgement remains hypothetical. No, because in practice what matters is not to identify an object correctly in every possible situation, but to be able to distinguish it from other objects for which it could be mistaken. The latter ability is essentially 'finite' however. Here is a simplified argument to that effect. Suppose that in a given context we already know that an object ξ must satisfy $A(x) \vee B(x)$, where A, B are mutually disjoint predicates on the inverse limit $\mathcal{M}$. If ξ does not satisfy A on $\mathcal{M}$, then there must exist a 'finite' stage s such that for $t > s$, $\mathcal{M}_t \not\models A(\xi_t)$. Hence for these t, and by condition (ρ) in fact for all t, $\mathcal{M}_t \models B(\xi_t)$, so that ξ satisfies B on $\mathcal{M}$. Since the formalisation of perception reports needs some additional technical machinery, this argument has to be elaborated, but the result is the same: in order to visually distinguish objects from each other, when the context provides the possible alternatives, a 'finite' amount of information suffices.

0.3.2 Decomposing models

The mathematically inclined reader may have wondered how 'special' models obtained as inverse limits of finite models are. For example, can every model be obtained this way? The following theorem goes someway toward a characterisation of the relevant class of models. First we need a definition.

Definition 3 *A model $\mathcal{M}$ is topological if there exists a compact Hausdorff topology on the universe of $\mathcal{M}$ such that all predicates are closed in the product topology.*

Theorem 2 *(W. Taylor [16]) Let $\mathcal{M}$ be a topological model, then there exists a inverse system $\langle \mathcal{M}_s \rangle_{s \in T}$ such that each $\mathcal{M}_s$ is a closed subset of some $[0,1]^k$, k finite, and $\mathcal{M}$ is the inverse limit of $\langle \mathcal{M}_s \rangle_{s \in T}$. In addition, if $\mathcal{M}$ is totally disconnected, then the $\mathcal{M}_s$ can be taken to be finite.*

The proof of the theorem is fairly simple. The hypothesis implies that the universe $|\mathcal{M}|$ can be embedded homeomorphically into a Hilbert cube $[0,1]^I$, or into a Tychonov cube $\{0,1\}^I$ in the totally disconnected case. Now let T be the set of finite subsets of I.

The usefulness of the theorem is limited however by the existential character of the hypothesis that $\mathcal{M}$ is topological.

0.4 Conditional quantification

In order to formalise perception we need one more ingredient. Recall our discussion of Miller and Johnson-Laird's formalisation: SEE is characterised in terms of a predicate PERCEIVE(x,y), which holds between a person x and a percept y; SEE(x,$A(y)$) is defined as PERCEIVE(x,y) $\wedge T(y)$, where T indicates the satisfaction of perceptual tests correlated with A. Miller and Johnson-Laird want to argue that the referential opacity of perception reports shows itself in the nonvalidity of quantifier exportation:

> [. . .] the logic of 'He sees a fish' thus both referentially and intensionally opaque. This logic is made explicit in [SEE(he, $\exists x$Fish(x))], a formulation that is compatible with hallucinations or with misperceptions such as $\exists y$ Shoe(y) $\wedge$ [SEE(he,Fish(y))](Miller and Johnson-Laird [14, P. 586]).

Although we share their intuition, it is clear that the proposed formalisation cannot be right, since the y inside scope of the SEE operator is supposed to denote a percept, while the same y outside the scope denotes a real object. Therefore a SEE statement must somehow involve both the unknown object and the percept derived from it. In this section we shall introduce the necessary machinery to do so: we want to view the process leading from an object to a percept as a form of 'filtering' The word 'filter' here should not be taken in its usual logical meaning (as a set of sets closed under intersection and supersets); it derives rather from physical analogues such as the Gaussian filters of Marr [13, 54 *passim*]. Their function is to blot out details which occur at some specified scale (hence at smaller scales). When applied to a picture, this type of filter introduces a blurring of the picture. Put differently, the effect is that pictures which differ only at scales smaller than the specified level, are perceived as identical. Hence, informally at least, there is a connection between filters and equivalence relations.

Logically speaking, a filter in the sense intended here is a new kind of generalised quantifier, which applied to a formula of n free variables, in general yields a new formula, also in n free variables. Hence this notion of generalised quantification differs from the more traditional Mostowski - Lindström generalised quantification, which does bind variables. It will be seen however, that the new notion of quantification is generalised in the sense that ordinary $\exists x$ is a special case.

To explain the motivating example, we return to the inverse systems and inverse limits of the last section. Given a model $\mathcal{M}$ and an inverse system of finite models $\langle \mathcal{M}_s \rangle_{s \in T}$ such that $\mathcal{M}$ is the inverse limit of the $\mathcal{M}_s$, define an equivalence relation R_s on assignments on $\mathcal{M}$ by

$$R_s(f, g) \text{ iff for all formulas } \varphi\, \mathcal{M}_s \models \varphi[f_s] \Longleftrightarrow \mathcal{M}_s \models \varphi[g_s].$$

Here, f_s is the assignments on $\mathcal{M}_s$ defined by $f_s(y) := f(y)_s$. Now define quantifiers $\exists_s$ by

$$\mathcal{M} \models \exists_s\varphi[f] \Leftrightarrow \exists g(R_s(f,g) \wedge \mathcal{M} \models \varphi[g]).$$

One should think of $\exists_s\varphi(x)$ as the set of objects which can consistently be taken to satisfy $\varphi(x)$ given the information at stage s (encoded in the model $\mathcal{M}_s$). Since $\exists_s$ is defined by an equivalence relation, it satisfies the $S5$ properties. The interpretation of the following theorem is that the $\exists_s$ can act as filters which reproduce the $\mathcal{M}_s$ from $\mathcal{M}$. The full details can be found in van der Does and van Lambalgen [17], for our present purposes a loose formulation suffices.

Theorem 3 *Under a mild assumption, for predicates A, $\mathcal{M} \models \exists_s A[f]$ iff $\mathcal{M}_s \models \exists_s A[f_s]$.* [5]

Let us now consider the meaning of $\exists_s\varphi$ in the general case, without the 'mild assumption'. Again we refer to Miller and Johnson-Laird's idea that in order for SEE(he, $A(x)$) to be true, x should satisfy perceptual tests for A-hood; we may add: tests at a certain level of accuracy s. The set $A_s := \{f|\mathcal{M}_s \models A[f_s]\}$ is one such test, satisfying $A \subseteq A_s$, but actually any formula ψ such that $A \subseteq \{f|\mathcal{M}_s \models \psi[f_s]\}$ defines a test for A. Now consider

Lemma 1 *$\exists_s\varphi$ is the unique quantifier satisfying, for all sets $\{f|\mathcal{M}_s \models \psi[f_s]\}$,*

$$\varphi \subseteq \{f|\mathcal{M}_s \models \psi[f_s]\} \Longleftrightarrow \exists_s\varphi \subseteq \{f|\mathcal{M}_s \models \psi[f_s]\}.$$ [6]

Since by the preceding lemma the quantifier $\exists_s\varphi$ represents the set of objects, or rather assignments, which satisfy all perceptual tests for φ at accuracy s, we may as well turn Lemma 1 into a definition. This leads us to the notion of a *conditional quantifier.*

Definition 4 *Let $\mathcal{M}$ be a model, $\mathcal{F}$ the set of assignments on $\mathcal{M}$, $\mathcal{G}$ an algebra of subsets of $\mathcal{F}$. By a* quantifier conditional on $\mathcal{G}$*—notation: $\exists(\bullet|\mathcal{G})$—we mean a mapping which applied to a set $\varphi := \{f \in \mathcal{F}|\mathcal{M} \models \varphi[f]\}$ yields an element of $\mathcal{G}$ such that*

$$(*)\ \ \forall C \in \mathcal{G}[\varphi \subseteq C \iff \exists(\varphi|\mathcal{G}) \subseteq C]$$

It will be useful to think of the conditioning algebra $\mathcal{G}$ as a bounded *resource*, so that conditional quantification is a form of resource-bounded quantification. Concrete examples of such resources $\mathcal{G}$ are given by the algebras $\mathcal{G}_s$, the Boolean algebras over the sets $\{f \in \mathcal{F}|\mathcal{M}_s \models \varphi[f_s]\}$ contained in $\mathcal{F}$. Another example is furnished by the ordinary existential quantifier. Let $\mathcal{G}_x$ be the Boolean algebra

[5] The result holds for predicates only, for arbitrary formulas no neat relation between truth on $\mathcal{M}_s$ and filtered truth on $\mathcal{M}$ can be expected.

[6] Strictly speaking, this result holds only on ω_1-saturated models or on inverse limits of an inverse system of finite models.

of sets of assignments definable by formulas which do not involve x free; then we have $\exists x\varphi = \exists(\varphi|\mathcal{G}_x)$. In general, think of $\mathcal{G}$ as the collection of available tests; then $\exists(\varphi|\mathcal{G})$ represents the collection of assignments that pass all tests contained in $\mathcal{G}$ for φ. In probabilistic terms, $\exists(\varphi|\mathcal{G})$ is the best estimate of φ on the basis of the information available in $\mathcal{G}$; hence we require $\exists(\varphi|\mathcal{G}) \in \mathcal{G}$.[7] Note that a quantifier satisfying (*), when it exists, is unique. Indeed, the Galois correspondence (*) implies that $\exists(\varphi|\mathcal{G})$ must be defined as $\bigwedge\{C \in \mathcal{G}|\varphi \subseteq C\}$. For this reason, Definition 4 is not yet quite what we want, because there may be $\mathcal{G}$ and φ for which $\exists(\varphi|\mathcal{G}) \notin \mathcal{G}$; but it suffices for the following discussion. A detailed technical treatment will be found in van Lambalgen and van der Does [18] and van der Does and van Lambalgen [17] .

It can be shown that if the conditioning algebra $\mathcal{G}$ is of the form $\mathcal{G}_s$, the conditional quantifiers $\exists(\bullet|\mathcal{G})$ have the following properties

1. $\exists(\mathbf{0}|\mathcal{G}) = \mathbf{0}$, $\exists(\mathbf{1}|\mathcal{G}) = \mathbf{1}$;
2. $\varphi \subseteq \psi$ implies $\exists(\varphi|\mathcal{G}) \subseteq \exists(\psi|\mathcal{G})$ (monotonicity);
3. $\varphi \subseteq \exists(\varphi|\mathcal{G})$ (coarsening);
4. $\exists(\varphi \vee \psi|\mathcal{G}) = \exists(\varphi|\mathcal{G}) \vee \exists(\psi|\mathcal{G})$ (additivity);
5. $\exists(\varphi \wedge \psi|\mathcal{G}) = \exists(\varphi|\mathcal{G}) \wedge \psi$ where $\psi \in \mathcal{G}$ ('taking out what is known').

Note that (2) and (5) imply that $\exists(\bullet|\mathcal{G})$ is the identity on $\mathcal{G}$

$$\exists(\varphi|\mathcal{G}) = \exists(\mathbf{1} \wedge \varphi|\mathcal{G}) = \exists(\mathbf{1}|\mathcal{G}) \wedge \varphi = \mathbf{1} \wedge \varphi = \varphi$$

for $\varphi \in \mathcal{G}$. Thus, (5) is the analogue of the Frobenius property in logic

$$\exists(\varphi \wedge \exists(\psi|\mathcal{G})|\mathcal{G}) = \exists(\varphi|\mathcal{G}) \wedge \exists(\psi|\mathcal{G}).$$

Interlude: seeing as filtering formalised With this machinery at our disposal, we now give a preliminary treatment of the logic of perception reports. The heart of the present approach is that a statement of the form 'I see a φ' is decomposed into two components.

1. The object that I see [x] satisfies all perceptual tests for being a φ *at the present approximation* [s].
2. I expect that this will continue to be the case when I move to more refined approximations [$t > s$].

By what has been said before, condition 1 can be rendered formally as

$$\mathcal{M} \models \exists(\varphi(x)|\mathcal{G}_s)[f].$$

Here, $\mathcal{M}$ represents the world, and f assigns an object in $\mathcal{M}$ to x. Since one has only $\varphi \subseteq \exists(\varphi(x)|\mathcal{G}_s)$, it does not follow that also $\mathcal{M} \models \varphi[f]$, which is as it should

[7] The connection between conditional quantification and conditional expectation in probability theory is spelled out in slightly greater detail invan Lambalgen and van der Does [18].

be given the hypothetical nature of perception. To take Miller and Johnson-Laird's example, it is perfectly well possible that $\exists x \exists (Fish(x)|\mathcal{G}_s) \wedge Shoe(x))$, even though Fish and Shoe are disjoint. Only when Fish is perceived completely accurately, i.e. when Fish $\in \mathcal{G}_s$, do we obtain referential transparancy, by 'taking out what is known' (property reftakingout above). Likewise with inferences of the form

I see a girl walk; therefore there is a girl whom I see walk.

Formally, this is the inference from

$$\mathcal{M} \models \exists (G(x) \wedge W(x)|\mathcal{G}_s)[f]$$

to

$$\mathcal{M} \models (G(x) \wedge \exists (W(x)|\mathcal{G}_s)[f];$$

Again, referential transparancy with respect to G is obtained only when $G \in \mathcal{G}_s$ with the help of 'taking out what is known'. (Condition 2 will be formalised as a default rule in section 0.5.1.).

That the conditioning algebra $\mathcal{G}$ is of the form $\mathcal{G}_s$ represents the situation in which the perceptual tests include all the tests for object recognition available at level s. What is interesting about conditional quantifiers is that we may play with the resource $\mathcal{G}$ to represent situations where there are fewer, or less accurate, perceptual tests available. This allows for the modelling of various forms of partiality relevant for perception. We shall discuss two examples. The first example concerns the influence of knowledge on perception. If I know that something is impossible, is it impossible for me to see it? No, and hallucinations such as Macbeth's dagger provide a case in point.[8] Formally we have the following situation. For a certain stage of approximation s: $\forall \overline{x}(\psi \to \neg\varphi)$, where $\psi \in \mathcal{G}_s$ (e.g. 'daggers don't float in front of one's face').
Does it follow that $\forall \overline{x}(\psi \to \neg\exists(\varphi|\mathcal{G}_s))$, i.e. that, even when given accurate information about daggers, I can't see a dagger float in front of my face? Intuitively not, but it does follow formally from the Galois condition in conjunction with the fact that the resource $\mathcal{G}_s$ is a Boolean algebra, so closed under negation. It would no longer follow if the resource would be generated by positive formulas only. The lattice of sets of assignments definable by positive formulas can be viewed as an example of a so-called *co-Heyting* lattice:

Definition 5 *A co-Heyting lattice (Lawvere [11]) is a lattice with $\vee$ and $\bigwedge$ such that*

$$a \vee \bigwedge_i b_i = \bigwedge_i (a \vee b_i).$$ [9]

.

[8] For a striking nonvisual example, compare phantom limbs.

[9] Note that the presence of infinite $\bigwedge$ guarantees that the conditional quantifier is always defined.

The second example of the use of conditional quantifiers is concerned with perspective. Recall that when defining inverse systems we allowed that predicates need not be defined on all models of the system. A concrete example where this is useful, is given by perspective: in a model which develops only the hind parts of a human, the predicate 'mouth' should not be applicable. In terms of perception reports, this means that a phrase like 'I can't see his mouth' should be construed as *denial*; this sentence is true when 'seeing his mouth' is 'undefined' because it is evaluated in the wrong perspective. Now it seems that our set up cannot model this, because $\exists(A|\mathcal{G}_s)$ is always defined, even when the predicate A is not interpreted on $\mathcal{M}_s$. Note, however, that in such a case, $\exists(A|\mathcal{G}_s)$ will typically equal $\mathbf{1}$. Since clearly one cannot consider the situation where $A \neq \mathbf{1}$ and $\exists(A|\mathcal{G}_s) = \mathbf{1}$ a situation of seeing A, it seems best to leave $\exists(A|\mathcal{G}_s)$ undefined in such cases.

To achieve this, the frames which determine the conditional quantifiers must be redefined so that they may lack a top element.

Definition 6 *A* pseudolattice *L is a partially ordered set in which meets and joins of finite non-empty sets exist. A pseudolattice L is an* evidential $\vee, \bigwedge$-frame *if it is closed under arbitrary non-empty meets, such that the following distributive law holds:*

$$a \vee \bigwedge_I b_i = \bigwedge_I (a \vee b_i).$$

Note that in lattices in which arbitrary finite meets and joins exists, top and bottom can be defined by $\top = \bigwedge \emptyset$ and $\bot = \bigvee \emptyset$; but a pseudolatice may lack top and/or bottom.

A typical example of an evidential frame can be obtained from a model $\mathcal{M}_s$ by taking the collection of sets of assignments definable by a positive formula, this time however *not* including top or bottom.[10] If there exist disjoint predicates on $\mathcal{M}_s$, then $\mathbf{0}$ will be in the evidential frame, otherwise not. If the universe of the model is not covered by a formula, in the sense that for some φ, $\{f|\mathcal{M}_s \models \varphi[f]\}$ equals the full set of assignments, then $\mathbf{1}$ will not be in the frame.

This construction allows us to represent the form of partiality which makes the following argument invalid:

> Whitehead saw Russell. Russell winked. Therefore Whitehead saw Russell wink.

This argument would be formalised by

$$\mathcal{M} \models \exists(x = r||G_s) \wedge \forall x(x = r \leftarrow W(x))[f] \text{implies} \mathcal{M} \models \exists(W(x)|\mathcal{G}_s)[f].$$

Here, r is a constant denoting a thread representing the individual 'Russell' in the inverse limit. In our original set up this would be a valid argument due to the monotonicity of the conditional quantifier. However, if we change the resource

[10] Also, we do not allow $=$ in the formulas occurring in the frame

$\mathcal{G}_s$ so that it is an evidential frame only without top, the argument becomes invalid in the sense that the premisses may be true while the conclusion lacks a truth value.[11]

0.5 Veridical perception

Recall that we took the expression 'I see a φ' to mean the conjunction of 1 and 2, where the stages s, t refer to some inverse system:

1. The object that I see $[x]$ satisfies all perceptual tests for being a φ *at the present approximation* $[s]$.

2. I expect that this will continue to be the case when I move to more refined approximations $[t > s]$.

This separation is dictated by the logic of perception, since all perception is approximate. The second condition is evidently non-monotonic: more precise information may contradict the expectation expressed in 2. As it stands, 2 does not yet express that our perception will be veridical, because veridicality is a claim about the world, not about our perceptual apparatus. To make the connection, we need a result which says that if φ is true in every approximation, then φ is true ('in reality', i.e., on the inverse limit). This can indeed be shown, so that combined with this result, 2 yields *veridicality*.

The purpose of this section is to give a precise formulation to veridicality conceived of as a defeasible rule. We first discuss the standard format for defaults, and then present a slightly deviant form, more suitable for applications to perception. This is then applied to study veridicality inferences.[12] For a detailed discussion for inferences involving logical constants we refer the reader to van der Does and van Lambalgen [17].

0.5.1 Pragmatic inference from default rules

In Reiter's version of default logic (Reiter [15]), a *default* is a rule of the form

$$\alpha : \beta_1, \ldots, \beta_n / \omega$$

where α is the *prerequisite* of the rule, $\beta_1, \ldots, \beta_n$ are its *justifications*, and ω is its *consequent*. The customary interpretation of the rule is: 'if α has been derived from the background knowledge and $\beta_1, \ldots, \beta_n$ are consistent with what has been derived, conclude ω'.

A *normal* default is one in which there is a single justification which is identical to the consequent; this is the kind of default of interest to us. Normally,

[11] In this case the set $\{f | f(x)_s = r\}$ need not be in $\mathcal{G}_s$, so that $\exists(x = r || G_s)$ may be coarser than that. This is as it should be, if we think of r_s as a full 3D model.

[12] The interpretation of default rules owes much to conversations with Frans Voorbraak.

defaults are used to express rules with exceptions, such as 'Birds fly', formalised as

$$B(a) : F(a)/F(a),$$

for every constant a. A default theory consists of a set of facts and a set of default rules. The facts ('Tweety is a bird') are taken to be specific and reliable information, and the defaults represent general information. In our case the situation is slightly different: the specific information consists of perceptual judgments, which are approximate, hence defeasible. The default rules should express the expectation that the judgment will continue to be true in more refined approximations. In this slightly different situation, the intended interpretation of defaults also undergoes a subtle change.

For the following discussion an inverse system $\langle \mathcal{M}_s, h_{st} \rangle_{s,t \in T}$ with inverse limit $\mathcal{M}$ and a corresponding system of quantifiers $\exists(\bullet|\mathcal{G}_s)$ is assumed to be given. If 's' stands for the present approximation, the statement 1

> with the present approximation, x, the object that I focus on, is identified as a φ

is formalised as $\exists(\varphi(x)|\mathcal{G}_s)$. In Reiter's format, statement 2 would then be expressed by the default rule

$$\exists(A(x)|\mathcal{G}_s) : \exists(A(x)|\mathcal{G}_t)/\exists(A(x)|\mathcal{G}_t) \text{ (for } t > s\text{)},$$

which says that if evidence at stage s derives $\exists(A(x)|\mathcal{G}_s)$, and if $\exists(A(x)|\mathcal{G}_t)$ is consistent with the evidence (where $t > s$), then assume $\exists(A(x)|\mathcal{G}_t)$

However, a different, 'evidential', interpretation better fits the perceptual situation. The discussion of Marr's theory of object recognition in 3.5 strongly suggests that the stage s, given by 1, should represent the *maximal available information*: we move downward in the hierarchy of 3-D models until the finite resolution of the image leads to a branching, hence to uncertainty. Suppose the visual evidence at stage s is summarised by $\exists(\psi|\mathcal{G}_s)$ where $\exists(\psi|\mathcal{G}_s) \neq \mathbf{0}$. One may think of $\exists(\psi|\mathcal{G}_s)$ as the description of a perceived scene at level s. We now make an additional observation $\exists(\varphi|\mathcal{G}_s)$. We may assume that $\exists(\psi|\mathcal{G}_s) \cap \exists(\varphi|\mathcal{G}_s) \neq \mathbf{0}$. By monotonicity $\exists\exists(\varphi|\mathcal{G}_t)|\mathcal{G}_s) \geq \exists(\varphi|\mathcal{G}_s)$. Now suppose that $\exists(\psi|\mathcal{G}_s) \cap \exists(\varphi|\mathcal{G}_t) = \mathbf{0}$. Then also $\exists(\exists(\psi|\mathcal{G}_s) \cap \exists(\varphi|\mathcal{G}_t)|\mathcal{G}_s) = \mathbf{0}$, and by Frobenius, $\exists(\psi|\mathcal{G}_s) \cap \exists(\exists(\varphi|\mathcal{G}_t)|\mathcal{G}_s) = \mathbf{0}$; whence $\exists(\psi|\mathcal{G}_s) \cap \exists(\varphi|\mathcal{G}_s) = \mathbf{0}$, a contradiction. It is crucially important for this argument that the perception $\exists(\varphi|\mathcal{G}_s)$ is performed at level s. If the perception were less accurate, say $\exists(\varphi|\mathcal{G}_r)$ for $r < s$, then the assumption $\exists(\psi|\mathcal{G}_s) \cap \exists(\varphi|\mathcal{G}_r) \neq \mathbf{0}$ does not contradict $\exists(\psi|\mathcal{G}_s) \cap \exists(\varphi|\mathcal{G}_s) = \mathbf{0}$. In this sense the perception has to be maximally informative relative to the background knowledge.

Motivated by these considerations we shall take a default to be a rule of the form

$$\exists(\varphi|\mathcal{G}_s)/\exists(\varphi|\mathcal{G}_t)$$

with φ a *positive* formula.[13] The reason for the restriction to positive formulas will become apparent in a moment. This rule should be interpreted as: 'if I

[13] A formula is positive if it is built from predicates using only $\forall, \exists, \wedge, \vee$.

have observed φ at stage s, and s represents the maximum available accuracy, then I may assume that I will observe φ at stage t.[14]

Having introduced the expectation inherent in every perceptual experience in the form of a default rule, we return to the principle of veridicality, which would sanction inferences of the form

¿From 'I see a φ' infer 'There is a φ'.

or formally

¿From $\mathcal{M} \models \exists(A(x)|\mathcal{G}_s)[f]$ infer $\mathcal{M} \models \exists x A(x)$.

To see the connection between the default rules and veridicality, assume that we have an assignment f such that for all t, $f \in \exists(A(x)|\mathcal{G}_t)$; does it then follow that $\mathcal{M} \models A(x)[f]$? For positive formulas, the answer is 'yes'.[15]

Theorem 4 *Assume the bonding mappings are surjective. For* positive *φ,*

$$\bigcap_{s \in T} \exists(\varphi|\mathcal{G}_s) = \varphi.$$

This result fails already for negations of atomic formulas! □

A related result, with 'positive' replaced by 'positive primitive', holds if the bonding mappings are not necessarily surjective. Thus we have established that veridicality is a defeasible consequence of the expectation inherent in all visual perception.

0.6 Evidence for structural partiality

In chapter 7 of his book *Vision* [13, p. 357 *passim*]), Marr argues that his hierarchy of 3-D models may serve as a paradigm for semantics generally. He asks himself

> *What do you feel are the most promising approaches to semantics?*

and answers

> Probably what I call the problem of multiple descriptions of objects and the resolution of the problems of reference that multiple descriptions introduce. [...] I expect that at the heart of our understanding of intelligence will lie at least one and probably several important principles about organizing and representing knowledge that in some sense capture what is important about our intellectual capabilities. [...]

[14] The difference between Reiter's interpretation and ours is that we take the consistency of the justification to be relative to a stage s, whereas in Reiter's case it refers to an extension of the default theory. Here, we shall forego a discussion of the possible notions of extensions applicable in this context; see [19].

[15] The first theorems of this kind were given by [20].

1. The perception of an event or object must include the simultaneous computation of several different descriptions of it, that capture diverse aspects of the use, purpose, or circumstances of the event or object.
2. That the various descriptions referred to in 1. include coarse versions as well as fine ones. These coarse descriptions are a vital link in choosing the appropriate overall scenarios [...] and in establishing correctly the rôles played by the objects and actions that caused those scenarios to be chosen.

To give a simple example, if I am in the jungle and I observe a large writhing shape in front of me, I need relatively little detailed shape information to decide whether the object is a snake or a human, and to adjust my behaviour accordingly. On the other hand, to choose an appropriate course of behaviour during a faculty meeting may require close scrutiny of the facial expressions of the other participants.

Examples such as this suggest that there exist two forms of partiality in cognitive structure, which may be called 'truth conditional' and 'structural' partiality. The first form of partiality is familiar from situation semantics. From

John saw Mary swim.

it does not follow that

John saw Mary swim and Bill surf or not surf.

because the scene supporting 'John saw Mary swim' may not contain the predicate 'surf', even though it contains Bill. In this case, although we may have a multitude of scenes, objects are still taken to be pointlike, they have no internal structure.

This is different for the case of structural partiality. Here, we want to think of objects as coming with a specified granularity[16], so that we have the possibility to consider the same object simultaneously at different scales.

An obvious question is whether there really is such a big difference between the two forms of partiality: can't one always represent structure by means of suitable additional objects and relations, perhaps even numerical? Probably, but that is not quite the point.

We are interested in cognitive structure, and how it relates to semantics. In the context of a semantics of 'see', it is important how we are able to talk about what we see. This ability involves object recognition, attaching a name to the object, and pronouncing the name. The standard view of object recognition is that it requires matching with a stored 3D template.[17] Accordingly, we must distinguish between three stored representations: structural representation (describing the visual form of the object) semantic representation and

[16] One of the first papers emphasising the importance of this notion is Hobbs [8].

[17] There do exist alternative approaches, for example the work of Bülthoff and his colleagues in Tübingen; cf. Bülthoff [2]

phonological representation. We shall have more to say on the precise nature of this trichotomy below, but for now we wish to stress the importance of keeping the information in the structural description separate from the information in the semantic description, also for a model theoretic semantics. For instance, it is often said, following Talmy, that spatial prepositions require very little geometric structure both for the figure and the ground. This statement will be qualified below, using recent work of Herskovits, but it should be clear that it can only be discussed formally in a framework where objects can be represented in varying amounts of detail, with mappings that serve to keep track of the same object specified in different granularities. We shall now discuss these issues in more detail.

0.6.1 Locative prepositions

In several ways spatial prepositions are relevant to our concerns. In so far as they provide a window on spatial cognition, their semantics seems to indicate that space is not mentally represented as simply a collection of partial veridical 3D maps. Furthermore, closer examination of this semantics shows that a fundamental role is played by structure-erasing functions, much like the projections from an inverse limit to one of its approximating models, and that in general inverse systems and limits provide a useful language in which to discuss semantics of prepositions.

As regards the first point, let us briefly consider some subjective ways in which space is represented. *Cognitive maps* structure space in such a way, that navigating is facilitated: they encode landmarks and the connecting routes. I once read in an English hiking guide: 'This map is not drawn to scale. The easy parts have been reduced in size.' This is funny precisely because one man's cognitive map is another man's perdition, but it emphasises the fact that cognitive maps are distorted images of the world, presenting one part in greater detail than others. For another example, consider the spatial relation 'near'. Viewed purely spatially, 'near' is symmetric; however, it does not appear to be so represented since one cannot say

> The house is near the bicycle.

although one can say

> The bicycle is near the house.

The only way one can felicitously use 'near' is apparently in sentences of the form

> *Object* near *landmark*.

Together, these examples show that spatial representation is not a partial submodel of three dimensional space with the usual geometric relations, but rather a homomorphic image of the world containing predicates underdetermined by geometry.

Let us now turn to a more detailed investigation of the semantics of prepositions. The general scheme that Herskovits proposes is the following: first a number of structure-erasing functions are applied to figure and ground, and then a predicate is computed on the results. For example, in order to verify the truth of the sentence

> Jane walked across the streaming crowd.

the structure of the Ground ('crowd')is reduced by first enclosing the crowd in a single volume, then projecting the volume on the ground plane to obtain an area, and finally assigning an average direction of motion to this area. Likewise, the structure present in the phrase 'Jane walked' is reduced by considering only the path traced by the point Jane. The relation 'across' now applies if the direction of the path is orthogonal to the direction of motion of the area. There are several noteworthy features of this example. First, the preposition 'across' takes as arguments rather impoverished structures, namely a path and a plane surface with intrinsic directionality. Secondly, these impoverished structures differ in the amount of information that has been deleted. Thirdly, the impoverished structures were obtained by applying functions to Figure and Ground with clear geometric interpretation.

Herskovits [7] provides an interesting catalogue of such functions, called *object geometry selection functions*

1. Idealisations to point, line, plane, surface, ribbon ...
2. Gestalt processes such as 3D grouping (enclosure in a volume), completed enclosure, normalised shape, ...
3. Selection of axes and directions, for example model axis, principal reference axis, ...
4. Projections, for example on the ground plane, ...
5. Part selection, for example 3D part, free top surface, ...

We have seen instances of the first four types of functions at work in our example; the last function is illustrated by the sentences

> The tablecloth lay over the table.

> The cat lay under the table.

Here, the table is identified with the tabletop.

The point of all this is to show that the semantics of natural language actually requires the simultaneous representation of an object at different degrees of granularity. This *schematisation* is described by Herskovits [7, p. 169] as 'a process that reduces a physical scene, with all its richness of detail, to a sparse and sketchy semantic content', and by Talmy as follows 'a process that involves the systematic selection of certain aspects of a referent scene to represent the whole, disregarding the remaining aspects'. (Quoted in Herskovits [7, p. 169]

Formally, schematisation can clearly be modelled by a set of homomorpisms, since if f is an object geometry selection function, then for any object a, $f(a)$ should go proxy for a (cf. the quotation from Talmy), hence should have the same properties as a. In this context it is useful to allow partial homomorphisms as well, as will be seen in a moment. The set of schematising (partial) homomorphisms should be closed under certain operations. For instance given partial homomorphisms f, g defined on a model $\mathcal{M}$, with disjoint domains, one would like to be able to form the disjoint sum $(f \dot{+} g)$, such that (roughly speaking) $(f \dot{+} g)$ yields a two-sorted model, one sort corresponding to the decrease in structure given by f, the other corresponding to the decrease in structure given by g. In this way one can take account of the fact that a preposition takes arguments of different granularities. If we furthermore make the reasonable assumption that the product of two such schematising functions is again schematising, then the collection of schematic images of a given model (the 'real world') actually forms an inverse system. A preposition is a relation that lives on models $\mathcal{M}_s$ for s larger than some given t, but it differs from predicates representing nouns in that the latter, in principle depending upon very precise shape information, will decrease as s gets larger, whereas this is not so for prepositions. Once a preposition P is is interpreted on a model $\mathcal{M}_s$ by P_s its interpretation P_t on the model $\mathcal{M}_t$, where $t > s$, is given by $h_{ts}^{-1} P_s$. This reflects the fact that more refined information about the crowd will never lead to a falsification of the sentence 'Jane walked across the streaming crowd'.

Of course, the description just given provides only an abstract framework; a detailed formal investigation of prepositions would have to specify exactly the nature of the models $\mathcal{M}_s$ and the homomorphisms between them. For example, in the case of a Gestalt function such as 3D grouping, one would need a high dimensional vector space, and an estimation function which encloses a set of vectors in the space in the smallest volume specifiable by a fixed small number of parameters. The point of this excursion into prepositions was rather to emphasise that the notion of structural partiality, apparently necessary for a semantics of perception reports, occurs naturally in this branch of semantics as well.

0.6.2 Visual agnosia and cascade models

Visual agnosia, more precisely associative visual agnosia, is a disorder of the higher visual functions characterised by three conditions

1. difficulty in recognising a variety of visually presented objects, for instance in naming tasks

2. normal recognition of objects through other sense modalities such as touch

3. apparently intact visual perception in so far as relevant for object recognition.[18]

[18] Associative agnosia is opposed to apperceptive agnosia, where also lower visual functions

Visual agnosia, although rare, provides a glimpse of our cognitive architecture and it can be used as a guide to set up controlled experiments.

Recall that picture naming requires access to three different types of representation: structural knowledge about objects, semantic knowledge and phonological descriptions. Humphreys, Riddoch and Quinlan [9] studied the process of picture naming, first with a subject suffering from agnosia and then with normal subjects, with the aim of obtaining more detailed information about the interaction of these levels. If a subject is impaired at naming visually presented objects, the problem may reside either in one of the representations, or in accessing the representations. Their subject, JB, though severely impaired at picture naming, apparently had roughly normal structural, semantic and phonological representations. [19] If the representations are intact, the problem must reside in accessing representations. One may entertain two different theories on the exact nature of accessing these representations:

1. The process is *discrete* in the sense that information is only transmitted to the next stage after the construction of the representation has been finished; for example, the structural description of a picture or an object must be finished to the extent that no other description remains activated, before it is passed on to the semantic level.

2. On the other hand, the process could be a *cascade* in the sense that semantic information about a picture can be activated prior to the completion of the structural description of the object.

For us it is of interest that there exists a clear model theoretic distinction between the two views: on the first view, a semantic system is best represented as an ordinary first order model, with predicates applicable to objects whose structural description is completed, so, one might as well say, to unstructured objects; whereas on the second view, predicates should also be applicable when the structural description is not yet completed, so that it becomes important to keep track of the stages of structural description of an object. The latter option is more like an inverse system of first order models. Hence we view the experiment to be described as a rough indication of which type of semantic organisation is to be preferred.

Now suppose that, as in JB's case, the access of the phonological representation from the semantic representation is intact. In order to decide between the two theories, one may observe that in the discrete case, if accees from structural to semantic representation were disrupted, one would expect a uniform impairment, in the sense that there would not be significant differences across categories. JB's impairment, however, was more pronounced for categories with

are impaired. It is hard to draw exact boundaries here. On this topic, see Farah [6] from which the above discussion was taken.

[19] We refer to Humphreys et al. [9] for a discussion of the tests used in ascertaining this. It should be added though that Farah [6] argues that associative agnostics do not have intact perception after all. If true, that would cast some doubt on the following case description, though not on the experiment with normal subjects.

structurally similar exemplars (such as BIRD), than for categories with structurally dissimilar exemplars (e.g. FURNITURE). The cascade model can explain why this should be so. Essential to this model is that before the structural descriptions have settled down the corresponding semantic descriptions are activated. In the case of structural similarity, many structural descriptions will be activated simultaneously, hence also many semantic descriptions. This however greatly increases the possibility of error.

Humphreys et al. [9] devised the following experiment on normal subjects to decide between the discrete and the cascade theories. As we have seen, the discrete theory predicts that structural similarity or dissimilarity between pictures will have no influence on the probability that the subject will come up with the correct name, since name giving starts only after the structural description has been completed, even when this takes a relatively long time (as in the case of structurally similar pictures). The cascade theory, on the other hand, predicts that structural similarity between pictures must have an influence on the probability of a correct answer: before the structural description has stabilised, there is ample time for interaction between semantic and structural description.

More precisely, Humphreys et al. studied the interaction between picture name frequency and structural similarity of pictures. Name frequency (the frequency with which a name occurs in print) is thought to affect the access to a picture's phonological representation, hence should be conditionally independent of structural similarity (given the semantic representation). The experimental results showed that there is little effect of name frequency in the case of structurally similar pictures (whose descriptions take a fairly long time period to access), but a large effect in the case of structurally dissimilar pictures (which are relatively easy to access). This result is what the cascade theory would predict: since name information is made available during the completion of the structural description, name frequency, which pertains to the phonological representation, has no effect. A further interesting result was, that in the case of structurally similar pictures, the reaction times for naming correlated strongly with the degree of structural similarity, and not so for the case of structurally dissimilar pictures. This seems to show that there must be a relatively high degree of structural similarity before it starts affecting naming performance. One explanation for this phenomenon is that in the case of structurally similar pictures a superordinate, 'generic' structural description is activated, corresponding to a category name (say 'bird'; here the authors refer to Marr's hierarchy of 3-D models), which in turn activates descriptions of many exemplars belonging to the category, thus further slowing down the process of name-giving.

0.7 The picture that says it all

We have been concerned with the question how the logic of perception influences the logic of perception reports. Though this question is interesting in itself, it also points to an underlying theoretical issue: what is the proper way to approach natural language semantics? In particular, what do semantic representation stand for? Traditionally, one distinguishes between model theoretic semantics, where semantic representations stand for aspects of the world (or possible worlds) and conceptual or cognitive semantics, where semantic representations are mental objects. The latter view tends to belittle the importance of logic and model theory, apparently on the supposition that model theory entails a realist commitment.

Clearly the logic of perception reports is relevant to this issue. Suppose we have come to a consensus on what this logic is. If it is best explained by assuming that perception is direct pick-up in the Gibsonian tradition, then at least for the purpose of modelling perception reports, model theoretic semantics (in the form of situation semantics) is fine. On the other hand, if it is best explained by assuming that perception involves inference and construction, then cognitive semantics seems more appropriate. We have adopted the latter view, but we clearly see no obstacles to using model theory in this context. Rather, model theoretic semantics should elaborated to the point where it takes into account not only the real world, but also the set of representations of this world that humans construct. Since perception provides a systematic relation between the two, in principle the relation between world and representation is susceptible to mathematical treatment. The way we like to think of this is that semantic interpretation is *factored* through conceptual structure, as in the following commutative diagram.

$$\begin{tikzcd}
 & \text{expression} \arrow[dl] \arrow[d] \\
\text{real world} & \text{conc. struct.} \arrow[l]
\end{tikzcd}$$

Figure 3: *Expression & meaning*

That is, expressions are interpreted directly only on the representations in conceptual structure, but indirectly on the world, due to the nature of the (perceptual) link. The way we have set up things, with homomorphisms, or more generally filters mediating between world and representations, is undoubtedly too simple. Hopefully, others will come with more realistic proposals, covering a wider range of phenomena.

Bibliography

[1] J. Barwise and J. Perry. *Situations and Attitudes.* The M.I.T. Press, Cambridge, Massachusetts, 1983.

[2] H.H. Bülthoff, S.Y. Edelman, and M.J. Tarr. How are three-dimensional objects represented in the brain? research report AI Memo No. 1479, MIT AI Lab, MIT, April 1994.

[3] C. C. Chang and H. J. Keisler. *Model Theory*, volume 73 of *Studies in Logic and the Foundations of Mathematics.* North-Holland, Amsterdam, third edition, 1990. First edition: 1973, second edition 1977.

[4] F.I. Dretske. *Knowledge and the flow of information.* The M.I.T. Press, Cambridge, Ma., 1981.

[5] R. Engelking. *General Topology*, volume 6 of *Sigma Series in Pure Mathematics.* Heldermann, Berlin, 1989. Revised and completed edition.

[6] M.I. Farah. *Visual agnosia.* The M.I.T. Press, Cambridge, Ma., 1990.

[7] A. Herskovits. Language, spatial cognition and vision. In O. Stock, editor, *Spatial and Temporal Reasoning*, pages 155–202. D. Reidel Publishing Company, Dordrecht, 1997.

[8] J.R. Hobbs. Granularity. In *Proceedings IJCAI85*, pages 1–4. AAAI, 1985.

[9] G. W. Humphreys, M. J. Riddoch, and P. T. Quinlan. Cascade processes in picture identification. *Cognitive Neuropsychology*, 6:67–103, 1988.

[10] R.S. Jackendoff. *Semantics and Cognition.* MIT Press, 1983.

[11] F.W. Lawvere. Intrinsic co-Heyting boundaries and the Leibniz rule in certain toposes. In A. Carboni, M.C. Pedicchio, and G. Rosolini, editors, *Category Theory*, pages 267–297. Springer-Verlag, 1991. Lecture Notes in Mathematics 1488.

[12] D. Lewis. General semantics. In D. Davidson and G. Harman, editors, *Semantics of Natural Language*, pages 169–218. D. Reidel Publishing Company, Dordrecht, 1972.

[13] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* W.H. Freeman, San Fransisco, 1982.

[14] G.A. Miller and P.N. Johnson-Laird. *Language and Perception.* Harvard University Press, Harvard, 1976.

[15] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[16] W. Taylor. Some constructions of compact algebras. *Annals of Mathematical Logic*, 3(4):395–437, 1971.

[17] J. van der Does and M. van Lambalgen. A logic of vision. Research report LP–97–07, ILLC, University of Amsterdam, 1997. 80 pp. To appear in *Linguistics and Philosophy*, 1999.

[18] M. van Lambalgen and J. van der Does. A logic of vision. Research report LP–96–14, ILLC, University of Amsterdam, 1996.

[19] F. Voorbraak. A nonmonotonic observation logic. Research report LP–97–09, ILLC, University of Amsterdam, 1997. Submitted.

[20] F. B. Wright. Convergence of quantifiers and martingales. *Pacific Journal of Mathematics*, 6:296–307, 1962.