



UvA-DARE (Digital Academic Repository)

Sculpting the space of actions: explaining human action by integrating intentions and mechanisms

Keestra, M.

Publication date
2014

[Link to publication](#)

Citation for published version (APA):

Keestra, M. (2014). *Sculpting the space of actions: explaining human action by integrating intentions and mechanisms*. [Thesis, fully internal, Universiteit van Amsterdam]. Institute for Logic, Language and Computation.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

3 DAVID MARR AND THE INVOLVEMENT OF CONCEPTS IN MULTI-LEVEL EXPLANATIONS

When we hear someone making vocal sounds, we still may ask whether or not that person is singing. Given variabilities between persons, generations, cultures, and so on, the behavior may need to be interpreted or defined otherwise than singing. Obviously, most aspects of the vocal expressions and causal conditions will remain the same, irrespective of our recognizing it as a case of singing. So the ambiguity in our classifying it as ‘singing’ may not be such that we mistake it for different functions like writing, gesturing, or other expressive actions: the voice must be involved and the sounds must be intended to having some expressiveness in order to qualify as singing. However, there are many different techniques of singing, which do not all employ our vocal tract in the same way or require the same cognitive processes of keeping melody and rhythm, or harmonizing with other instruments or voices, for example. Furthermore, whether or not vocal sounds are fulfilling a particular function is not easy to decide: we would in some cases accept vocal sounds as singing if they did not have an expressive function, for example, even though perhaps in most situations – as suggested from an evolutionary perspective – singing does fulfill a particular function. To grasp the function of singing we must usually consider the context in which the vocal sounds are being made, but in many cases we don’t need information about the context to decide whether someone is singing.

In other words, there are many different sorts of information that can be brought to bear upon someone’s singing, which can all be employed to decide whether that person is singing. If we want to explain that person’s behavior – whether or not it is singing – we will most probably need many different kinds of information. As noted above, in the life sciences such pluralism abounds. This has led to a situation in which biologists have to recognize the only limited significance of a particular theory, allowing room for the involvement of other theories on the same phenomenon (Beatty 1997). Theoretical integration is then a legitimate explanatory goal, while unification by way of a ‘single model of multiple causal factors’ is not, since: “contingency, context sensitivity, and nonlinear interaction among contributing causes preclude the success of these types of unification” (Mitchell and Dietrich 2006 78). The conclusion that Mitchell & Dietrich draw is that biological phenomena allow analysis and explanation at several levels of analysis and preclude unification. Obviously, in such a situation, reduction of a complex function to a particular level is even less an option.

In the previous section we discussed the arguments of Bennett and Hacker against

a particular form of reductionism: a reductionism according to which psychological functions are directly ascribed to the brain or particular parts of the brain (Bennett and Hacker 2003). They defend their position with a particular view on the nature and origin of the concepts that we use for psychological functions. On those grounds they emphasize that neuroscientists should recognize the fundamental task of preliminary conceptual analysis for any empirical study of the functions associated with these concepts or their neural correlates. What is lacking in their particularly critical account is a proposal contributing to a more fruitful relation between such a preliminary, conceptual analysis and the empirical studies aiming to clarify the neural correlates of the investigated function. For instance, as we suggested above and in (Keestra and Cowley 2011), the results of such analysis should not just function as a barrier to nonsensical judgments, but rather provide a heuristics to suggest novel investigations. As we will see in the next chapters of this part, there are other approaches possible that agree in the rejection of reductionism and accept the mereological principle, while differing from Bennett & Hacker with their strict separation of conceptual analysis and empirical investigations. These approaches explicitly leave room for the causal and theoretical pluralism that seems appropriate for this subject. The first one that we will be discussing has had a large ‘inspirational influence’ in the field (Glennerster 2007) over the past decades. It may not come as a surprise that it was established by one of the authors explicitly reproached by Bennett & Hacker: David Marr.³⁴

3.1 The analysis of computations or tasks – not concepts - should guide scientific investigations

Conceptual analysis of ‘singing’ can yield a description of singing as a particular type of vocal sounds, often consisting of words set to melodies, often expressing a particular intention or mood. Such an analysis does offer limited information about the function of singing itself, nor is it particularly helpful in determining scientific investigations – other than denying that scientists are in fact observing a singing person, when his behavior does not consist of expressing vocal sounds, for example. An alternative analysis would not so much analyze the concept –although that will remain an important step – but would analyze the function to which it applies by

³⁴ They cite Marr writing that “our brains must somehow be capable of *representing... information...*” and subsequently criticize such verbs as representing information, decision making and the like to the brain (Bennett and Hacker 2003 70), which suggests indeed that Marr and others have neglected the mereological principle. Marr’s multi-level explanations allow a loose interdependency between such forms of analysis, precluding such mereological reasoning. Instead, Marr’s approach invites the heuristic use of analytic results we defended above and earlier (Keestra and Cowley 2011).

asking questions like: what kind of a task is singing, and are there subtasks that we can distinguish? For example, during song a person is coordinating semantic and grammatical knowledge in correspondence with tonal, dynamical and rhythmic components. This involves various cognitive tasks, but also an increased demand of motor control, determining the tension of the vocal chords, breathing behavior, and so on. A task analysis that helps to distinguish these components or sub-tasks that – in coordination with some other tasks probably - make up the task of singing enables researchers to investigate such a complex task that might otherwise remain unmanageable. Notice that such a task analysis is different from, though perhaps assisted by, a conceptual analysis, even though it can equally guide and constrain empirical research of singing. Marr has contributed importantly by arguing for the relevance of such a task analysis and for describing its relation to other types of investigations in the study and explanation of vision.

Marr elaborated his methodology while working on vision or visual perception. Notwithstanding the fact that vision has traditionally been a most promising subject for cognitive neuroscience and showed quite some progress, what Marr found to be lacking in this field was a delineation of what the object of vision research in fact is. For a while, he was impressed by successes from the so-called ‘feature detector’ approach in vision science, primarily aimed towards discovering particular brain cells that respond to specific features in a visual scene. An example is the successful discovery of the bug-detector in the frog retina and some similar discoveries in its wake. Scientists then assumed, in the words of Barlow - quoted by Marr - that: “the activities of neurons, quite simply, are thought processes. This revolution stemmed from physiological work and makes us realize that the activity of each single neuron may play a significant role in perception. (p. 380)” (Marr 1982 13). Barlow believed that ‘a single neuron could ‘perform a much more complex and subtle task than had previously been thought’ and Marr confesses that: “[t]ruth, I also believed, was basically neural” (ib. 14). It is such reductionist belief that was countered in the previous sections, promoting a fundamental role for conceptual analysis for cognitive neuroscience. However, Marr developed another methodological response to it.

The hypothesis that explanations of visual perception could rely on the activities of so-called ‘grandmother cells’ and its kin, turned out to be fruitless (Marr 1982 15).³⁵ For one, it turned out that the number of such feature-detecting cells that

³⁵ Apart from the fact that it is implausible to be able to find such a grandmother cell if it were a single cell in the multi-billion cells, it is also principally impossible to falsify that the cell would not respond to any other face or figure. Indeed, it is more likely that instead of single ‘grandmother cells’, the brain contains ‘ensembles’ of cells that together represent a complex object, each cell responding to a different aspect of the stimulus (Gross 2002).

were discovered was extremely limited. More importantly, even if the ‘apocryphal’ grandmother cell were detected, Marr realized that crucial questions would remain like: “why or even how such a thing may be constructed from the outputs of previously discovered cells” (Marr 1982 15). Since we normally expect from explanations to clarify why or how certain facts hold, this particular lack of answers in the feature-detector approach dissatisfied Marr. Trying to make up for the lacuna, he proposed a methodology that allows combining the traditional results of neurophysiological and psychophysiological investigations of vision with answers to questions about the functional role of such facts. Such answers are presented in a so-called ‘computational theory’³⁶ and added to the other theoretical results that already figured prominently in the explanation of vision: results stemming from neurophysiological investigations and algorithmic descriptions of the neurophysiological properties and activities that were delivered by such investigations. In the next section we will discuss how Marr envisions this connection between different types of results in the explanation of a cognitive function like vision. For now, let us look more closely at what this computational theory amounts to – especially since it is this aspect of his methodology that Marr considers a particularly important contribution to the field (Marr 1982 330).

It is important to realize that a computational theory is concerned with a so-called competence³⁷ theory, only aiming at the formulation of the ends of a particular task, without consideration of the specific means to reach those ends (Marr 1977a).³⁸ As such, the computational theory focuses on theoretical ingredients that are not included in the feature-detector approach, since the latter is mainly interested in the implementation of a particular task. In contrast to Marr, Barlow and others believed to be able to deduce from evidence concerning the implementation as task its relevant functional properties, as can be seen from Barlow’s influential first dogma of neuroscience research: “A description of that activity of a single nerve cell which is transmitted to and influences other nerve cells and of a nerve cell’s response to

³⁶ Marr’s use of ‘computational’ has led to some controversy. For instance, it is different from that of Fodor, who is concerned more with ‘how’ veridical features of the environment are represented (Kitcher 1988).

³⁷ Marr himself refers to the computational theory as a competence theory (Marr 1977a), alluding to this term that Chomsky coined for the cognitive science of language, making a distinction between a competence and the way how such a competence is actually performed. As the term is here not without difficulties, I’ll make only limited use of it. As it seems that Chomsky has mistaken Marr’s computational theory for an exclusively internalist theory – about which more below – comparison of their ideas is beyond the scope of my present discussion. See (Silverberg 2006) for a defense of Marr’s computational theory against Chomsky’s interpretation of it.

³⁸ Elsewhere Marr made the comparison with the development of the theory of thermodynamics, which shows that a top-level theory may be useful even in the absence of: “a description in terms of mechanisms or elementary components” which appeared only afterwards (Marr and Poggio 1977 2).

such influences from other cells, is a complete enough description for functional understanding of the nervous system.” (Barlow 1972, 380, cited by Marr, 1982, 13) The computational theory that Marr envisaged, on the other hand, was clearly not a functional understanding that could simply be extrapolated from properties of single nerve cells. On the contrary, Marr argues that such a method easily overestimates the relevance of an understanding based upon neural cell properties alone.

Although it is interesting to discover feature-detectors in a visual system, only a computational theory can help us to realize that such detectors might be simply misled in practical reality. For example, a ‘bar-detector’ may be misled, since an edge of light may be mistaken for a bar if perceived by a single cell (Marr & Hildreth, 1980, 188). Indeed, it is reflection upon such potential flaws that emphasize the relevance of a computational theory when explaining a visual system. Such a theory might force the scientist to acknowledge the importance of ‘the discovery of valid constraints on the way the world is structured’ (Marr 1980). What these constraints amount to will be clarified below. First, a specification of the computational theory in more abstract terms is required.

Now the analysis of the task at hand, referred to by Marr as a ‘pure competence theory’ (Marr 1977a) or the computational theory, does not so much concern a logico-grammatical analysis of the meaning of the concept referring to that task – like it was demanded by the approach discussed in chapter I.2. Instead, a computational theory informs us with regard to a particular task about: “What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?” (Marr 1982 25, figure 1-4).³⁹ Nonetheless, a first answer of those rather abstract questions may be found in the case of vision when we reflect for a moment on the question what seeing in fact is: “The plain man’s answer (and Aristotle’s, too) would be, to know what is where by looking. In other words, vision is the *process* of discovering from images what is present in the world, and where it is” (Marr 1982 3, italics in original). According to this plain description, the goal of vision is: getting to know something about the world, and in particular what is there and where it is. This is a rather general description, perhaps most remarkable for the fact that it describes vision without explicitly relating it to other cognitive functions

³⁹ Bechtel even argues against Marr’s calling the upper level a ‘computational theory’: “The name is misleading since at this level the researcher is not concerned to explain the computational procedures, but rather to specify the task to be performed by the computation system, why that task is to be done and the constraints the task itself imposes on the performance of that task” (Bechtel 1994 4). It appears that Marr was at the time following the MIT convention of labelling any information processing task a computation, instead of a task or function. This convention was perhaps only superseded with the development at MIT of robots with a subsumption architecture, without a crucial role for a comprehensive representation, as in (Brooks 1991).

– a limitation that has provoked dissent, as we will see in section I.3.5 and in part II.

Leaving the latter issue aside, the computational theory still asks us questions that are more specific than the question that led to the ‘plain man’s answer’ regarding vision. Especially the aspects of appropriateness and of the logic of the strategy used for a computation merit further discussion. What these aspects imply can be illustrated with computational theories that concern specific subtasks of vision. Examples of components of vision for which a computational theory has been developed by Marr are the recognition of shapes from contours (Marr 1977a), stereopsis (Marr and Poggio 1979), and the detection of edges (Marr and Hildreth 1980). For instance, according to its computational theory, stereo vision requires that for two separate images – one for each eye – their symbolic descriptions are being matched, where the disparity between the images is being measured (Marr and Poggio 1979). How this is to be done and what neural implementation may be responsible for it is another matter. Relevant for the computational theory are still such ingredients that are relevant for “the logic of the strategy by which it can be carried out” (Marr 1982 25, figure 1-4) and which do not refer us to the performance specifics of the system or organism under investigation.

3.2 Constraints that co-determine the computations’ appropriateness

Marr’s analysis of the computational theory appears at closer reading to consist of two components. First, it offers a description of a particular task in rather general terms. Second, it then offers information on additional ingredients that make that computational theory plausibly effective. Both components are not of a particularly technical nature. For instance, he considers the paradoxical fact that when we perceive only figures’ black contours, as in Picasso’s ‘Rites of Spring’, these “tell us more than they should about the shape of the dark figures” (Marr 1977a). The competence to be explained is therefore how we can derive information about the objects when we are perceiving only two-dimensional black contours, sometimes even overlapping outlines of different objects. The explanation Marr offers depends upon two ingredients which necessarily underlie this capacity: “implicit in the way we interpret an occluding contour, there must lie some *a priori* assumptions that allow us to infer a shape from an outline” (Marr 1977a 441-442, italics in original) – where these assumptions concern the surface and structure of perceived objects. Such assumptions are more generally involved in cognitive tasks and should therefore be included in the computational theories underlying these tasks.

Marr uses several examples in his discussion of such assumptions. For example, the appropriateness and the logic of the computations that we perform with a cash register

in a shop rely upon “the rules we intuitively feel to be appropriate for combining the individual prices” and these rules “in fact define the mathematical operation of addition” (Marr 1982 22). The unexpected nature of such rules or assumptions stands out even more clearly when he discusses an airline reservation system that functions appropriately, depending on its capability of taking into account relevant properties of the world.⁴⁰ These properties must constrain the task if it is to provide results that are useful for the subject. Marr argues that an explanation of the computational theory of that system must not only refer to the properties of its computers, but also needs to include information about “what aircraft are and what they do; about geography, time zones, fares, exchange rates, and connections” and then expands that list even with “something about politics, diets, and the various other aspects of human nature that happen to be relevant to this particular task” (Marr 1982 5). This list of information may not yet give us much insight into a computer, nor does it yield insights in the specific ways in which the computer makes reservations, but the list is necessary to understand the kind of computations it carries out when processing reservations. Indeed, an airline reservation system that is not constrained by the presence of different time zones or local political unrest will yield results that are unrealistic or conflicting with other flights. The logic of its computations therefore depends partly upon its handling of these constraints. A feature-detector approach to the reservation system, without a computational theory of making reservations, would in this case not be able to explain why geographical differences or politics are processed by particular parts of the computer network. Consequently, it would be difficult to determine how detectors are related to environmental or internal parameters, to judge whether or not connections between detectors are functional, and so on.

One may believe that the inclusion of such constraints in a computational theory of a particular function makes the explanation of that function ever more complex. And to the extent that it forces the investigator to reflect upon a function’s environmental conditions and consider which conditions are potentially relevant, this is certainly true. On the other hand, such constraints facilitate the task of explaining why a particular computation’s strategy – being therefore only partly determined by its algorithmic and implementation theories, as we will see shortly – is indeed

⁴⁰ In this context, it has been discussed whether Marr’s theory is methodologically individualistic, or rather externalist in nature. The strict requirements that Morton poses on ‘non-solipsist’ accounts of vision or cognition, and which Marr’s approach does not meet, do not appear convincing to us. However, we must leave that topic here aside, but see (Kitcher 1988 ; Morton 1993 ; Silverberg 2006) for different positions in this discussion.

appropriate and adequate.⁴¹ For without any constraints provided by the environment a system for vision would have much more difficulty in recognizing ‘what is where’: if our environment were not occupied mostly by rigid objects that have spatial contiguity, for instance, visual recognition of objects would be quite difficult (cf. Marr 1982 209).⁴² Obviously, the visual system that operates in such an environment would require an explanation that includes a very different computational theory. In any case, when investigating the process of vision, researchers devote part of their efforts to discovering “additional constraints on the process that are imposed naturally and that limit the result sufficiently to allow a unique solution” (Marr 1982 104). These constraints are in fact “properties of the visible world that constrain the computational problem and make it well defined and solvable” (Poggio 1981 259). For these properties in turn determine what kind of information is needed to solve these problems and must therefore be produced by the system (Marr 1982). As a result, these constraints do not just contribute to the process of successful visual information processing, but at the same time also facilitate vision’s explanation.

Compare vision with our example of singing again: when investigating the singing of a lightweight bird in the sky we will look for very different sound producing body parts than if we focus on singing whales that live in a completely different medium and have different body properties. This example also shows what seems to be lacking in Marr’s approach: a consideration of the possibility that a competence may have different functions for different animals which may be partly dependent upon their other competences and even upon general properties of their existence and environments. Are the criteria for appropriateness of vision in humans not different from those in eagles or in rabbits? Even though it seems useful to focus on a particular competence in isolation from others, it may also risk leaving out important ingredients of an adequate explanation. The other approaches to be discussed in this part will take up this issue more explicitly.

Nonetheless, Marr’s requiring a computational theory in the explanations offered by cognitive neuroscientists has already expanded their methodology significantly.

⁴¹ Whether Marr’s approach leads in fact to an optimization theory with regard to both the targeted computational problems and the strategies used for solving those, opinions diverge. See e.g. the discussion between Kitcher and Gilman (Gilman 1994 ; Kitcher 1988). It seems assumed by Marr, indeed, that evolutionary selection processes have contributed to the development of a computation characterizable by a ‘logic of the strategy’ (cf. Marr 1982 105, 266). Dennett argues that behind both the optimalization and evolutionary aspects of Marr’s approach is his engineering stance (Dennett 1989 310-311), which seems to be a plausible hypothesis.

⁴² Although Marr believes that ecological psychologist Gibson came closest to developing a computational theory, he criticized Gibson for underestimating the information processing task of vision (1982, 29). However, Marr in turn underestimated the distinctness of Gibson’s approach in that it underlines the relation of vision to an animal’s other competences like action (Dennett 1989 310-311).

It has forced them not just to focus on the means by which certain competences are carried out, but also to first analyze those competences and their interaction with many other conditions. The question remains, however, how this computational theory has to be related to the other results of cognitive neuroscientific research. We will address this issue by focusing on Marr's strong defense of a multi-level explanation, which has meanwhile become prevalent in the field.

3.3 Two further levels for multi-level explanations

With the addition of a computational theory to the formulation of explanations of vision, Marr did not aim to cast aside the more traditional explanatory ingredients. Having learnt from ecological psychology to view: "the problem of perception as that of recovering from sensory information "valid" properties of the external world" (Marr 1982 29), he continued to explain vision in terms of visual information processing as being carried out by animal brains or other devices. As a result, the computational theory has to be considered as a part of a larger explanatory framework, since explaining vision requires several distinct theories simultaneously. Generally speaking, Marr concluded that: "[f]or the subject of vision, there is no single equation or view that explains everything. Each problem has to be addressed from several points of view – as a problem in representing information, as a computation capable of deriving that representation, and as a problem in the architecture of a computer capable of carrying out both things quickly and reliably" (Marr 1982 5). The other points of view to be taken into account are presented as distinct levels, to be added to a multi-level explanation. How these levels are interrelated and how they together make up an explanation will be discussed in section I.3.4. First, we will provide short accounts of the two other points of view, or levels, in two subsections on the algorithmic and the neural implementation levels.

3.3.1 The algorithmic level and the representation of information

Singing from sheet music does not come naturally. Since there are different forms of notation, it requires specific education. Depending on the instrument one plays or one's voice, within our Western tradition of music notation a musician must even learn to read in different keys and – if she plays organ - learn to read figured bass notation. Moreover, the hierarchical representation of musical information also plays a role in listening to music, as practiced listeners turn out to have different expectations than others (Justus and Bharucha 2001). Such cognitive representations develop over time and brain activation patterns together with behavioral responses show differences between individuals, ages, and even gender in the brain's musical

information processing (Koelsch, Grossmann et al. 2003). Independent of the contents of the auditory stimulus, the recognition and processing of information appear to differ widely. The representation of information has a great influence on these differences in musical competence, just as it does for vision.

Depending to a large extent on the constraints offered by the environment, the computational theory was found to focus on the ‘what’ and ‘why’ of a particular competence. The competence of perceiving objects in the world can be understood, according to Marr, as a “mapping from one kind of information to another” (Marr 1982 24). Given this information processing perspective, it is not surprising to find that: “[i]n the center is the choice of representation for the input and output and the algorithm to be used to transform one into the other” (Marr 1982 24). As the examples of recognizing objects in black contours or seeing depth from two different, two-dimensional retinal images illustrate, it is not just the gap between the environmental appearance and its perception by us that needs to be explained, but also how the information is represented in the retinal images and how their combination then allows stereopsis.

The focus of the algorithmic theory, present at the level between the computational and the neural implementation levels, is the rather abstract but crucial issue of how information gets represented and the related issue of the transformations applied to this represented information. The study of information representations and transformations is typically done in psychophysical research, employing response-time, delay, and interference tasks, or mental rotation tasks (Marr 1982 26) to develop hypotheses about “the scheme for a computation” (Marr 1980).⁴³ The relevance of such a scheme or representation of information is clear once one realizes how the computation of addition, for example, can be performed with Roman or with Arabic numbers or even by using one’s fingers. Each of those representation forms has its own merits and disadvantages and subjects may shift from one to another representation format as the occasion demands.

What is important to note, as Marr emphasized, is that the choice of a representation format is not just a matter of usefulness. In fact, “there is a trade-off; any particular representation makes certain information explicit at the expense of information that is pushed into the background and may be quite hard to recover” (Marr 1982 21). He illustrates this not just with the comparison of Roman and Arabic numbers, but also with the necessary conversion of a human decimal number system to the binary

⁴³ Until the work of Shepard and Metzler in 1971 visual psychologists had not recognized the importance of an algorithmic theory, according to Marr (Marr 1982 10).

representation format suitable for electronic devices. Addition and subtraction, recognition of powers of ten or equal numbers are for us much easier to carry out in a decimal system than in a binary one. Represented in a binary representation format, we would even have difficulty recognizing powers of ten, for instance. It should not be surprising, as we will indeed find out later, that differences in the implementation of computation tasks do affect the choice of the representation format. The centrality of this algorithmic level will become more evident in section I.3.4, where Marr's account of the relation between levels of explanations will be discussed.

There being several representation formats to choose from, the question arises how and on what grounds an engineer chooses one, or which format appears to have had most success in evolutionary history. Marr here makes a distinction between human vision and the visual information processes in animals. The feature-detectors that were successfully found in animals, like the frog's bug-detector and the rabbit's hawk-detector appeared to depend not just on particular hardware which may differ between animal species, but also on the particular algorithms that are performed with this hardware. A particular retinal image in combination with some other information immediately makes the animal perceive a potential prey or predator, allowing it to respond fast – even if at times it will attack or flee for an innocent object. Such representation errors are the cost that comes with the fact that each animal “can confidently be expected to use one or more representations that are nicely tailored to the owner's purposes” (Marr 1982 32).⁴⁴ Humans, however, are exceptional in this respect, Marr believes.

Unlike animals, humans do not just have ‘special-purpose mechanisms’ but human vision “seems to be very much more general” (Marr 1982 32). Given the wide range of information that humans can derive from their retinal images and the different computations that they perform on them, Marr argues that human vision must use representation formats – especially in the so-called early stages of visual information processing – that do not force information into the background in such a way that it is not recoverable for potential use in later processing.⁴⁵ Developing a plausible algorithmic theory for human vision should therefore obey the ‘principle

⁴⁴ Analyzing more in detail fly vision, Marr suggests that perhaps 60% of its very simple vision representation consists of three components that allow it to detect a suitable landing spot and a potential mate (Marr 1982 32-35), admitting for the rest very sparse information about the world.

⁴⁵ Even if Marr was right about the existence of these early vision sketches, it could still be the case that they are an “an accidental by-product of its function in guiding motion for the purpose of avoiding danger and securing food, shelter, and other objects of desire” (Hatfield 1991 177). The converse would also be possible: vision may most of the times be the result of the animal's active ‘probing’ of its environment, but that does not preclude the possibility of an by-product in the form of a general-purpose sketch – a possibility that would counter some arguments against Marr in (Noë 2004).

of least commitment': "one should never do something that may later have to be undone" (Marr 1976 485). Feature-detectors do not comply to this principle, as their informational focus is on very particular parts of information to the detriment of other information that could otherwise be used in later processes, potentially correcting earlier misperceptions.

This emphasis on the general purpose of human vision also made Marr deny an important role for a memorized database of objects, that many researchers at the time thought would enable quick recognition in humans. Since such recognition in the early stage of vision would depend on an algorithmic theory or representation format that would foreground particular features of information, it would potentially come at some cost with respect to other information features. This is related to the supposed 'modularity' of early vision, which we will address at the end of our discussion of Marr's work. Apart from that, the assumption of the general expediency of human vision as opposed to the more specific functionality of animal vision has met with serious criticism.⁴⁶ Be that as it may, the fact remains that representation format has implications for visual and other cognitive processing. More relevant to the present context, it underscores that for a comprehensive explanation of a cognitive function like vision we need to take into account several forms of insights, stemming from different scientific disciplines and integrate these in an interdisciplinary explanation. Insights that may at times be quite disparate, presenting a challenging task for those who strive to combine them.

3.3.2 The implementation level and neuroscientific evidence

In light of the current omnipresence of neuroscience results in contemporary debates about psychological functions and the far-reaching conclusions that are often drawn from these by neuroscientists and lay-people alike, it may come as a surprise that Marr's influential methodology includes neuroscientific results only as the lowest of three – or even four⁴⁷ – different levels of analysis and explanation. As argued above, it was his disappointment with the relevance of neuroscientific evidence regarding

⁴⁶ Especially so-called enactive views of perception underline that human vision is often noticeable geared towards the specific actions or sensori-motor relations that a subject undertakes (cf. Noë 2004 ; Thompson 1995). Perhaps more plausible is a recent proposal suggesting that vision consists of a mutual interaction between early vision and top-down influences of object recognition in a Bayesian inferencing mode (Yuille and Kersten 2006).

⁴⁷ The four-level account presented in (Marr 1980) included a next to lowest level, consisting of assemblies of the basic components and activities that make up the lowest level. This concurs with the recursive decomposition of an explanatory mechanism that can be carried out according to the mechanistic explanatory methodology that will be discussed below in chapter I.5. An alternative addition of a fourth level is level 1.5 by Peacocke. This amounts to the insertion of a level between the computational and the algorithmic level of such theories that yield functional equivalence classes without yet specifying algorithms (Peacocke 1986).

feature-detectors that convinced him of the importance of multi-level explanations in cognitive neuroscience. For some time, a simple mapping of mental states to brain states was attractive to him, too: “[a]t the time the eventual success of a reductionist approach seemed likely.” (Marr 1982 13). It was his realization that many neuroscientific results yield inadequate information about what contribution the neural activities make to the investigated function that disappointed him. Briefly stated, this led him to conclude that “reductionism does not imply constructionism” (Marr and Poggio 1977 2): it turned out to be impossible to derive from the computational theory of a cognitive function alone the information that would enable someone to build such a device, nor was it possible to assemble a functional system on the basis of evidence regarding potential neural components alone. For an adequate explanation of the perception of a visual scene, insights were necessary about the computations and the algorithms involved. However, this did not imply that insights about components were irrelevant. To begin with, in order to be performed, these obviously needed implementation in some form of hardware like an electronic device or a brain.

The lowest level of explanation concerns the ‘hardware implementation’ and has to answer: “[h]ow can the representation and algorithm be realized physically?” (Marr 1982 25, figure 1-4). This in itself is a complex issue, for to the same degree that algorithms differ between different animal species, different implementations are possible. To draw conclusions about particular implementations, investigations of the hardware of different species or other devices are required to provide information about: “how do transistors (or neurons) or diodes (or synapses) work?” (Marr 1980 199). The relevance of neuroscientific investigation for such information is unmistakable. However, at the same time some caution is in order.

The cash register discussed in section I.3.2 already suggested that addition and deduction can be performed in different ways, using different representation formats and algorithms. Such differences are not just important to consider for engineers designing a register, but also for neuroscientists investigating potential brain processes that can carry out those computations. Acknowledging this variability, Marr appears to subscribe to the assumption that cognitive tasks are *multiply realizable*, which implies that these tasks (or the ‘psychological predicates’ Putnam dealt with in his seminal paper (Putnam 1967)) not only allow description without recurrence to correlated brain-states but can also be correlated to multiple non-identical brain states.⁴⁸

⁴⁸ As Marr realizes that frogs and rabbits may have different implementations of comparable computations, his position implies subscription to the idea of local multiple realizability of some functions as described in (Kim 1992). With respect to this Wimsatt remarks that it is: “entertaining to see philosophers of psychology act as if this characteristic is a special property of the mental realm” given its prevalence in nature (Wimsatt 2007 217).

At times it seems that Marr subscribes to a functionalist position, which attaches relatively little importance to the implementation level of the explanation of a particular cognitive function. This may be derived from statements like: “trying to understand perception by studying only neurons is like trying to understand bird flight by studying only feathers. It just cannot be done” (Marr 1982 27). This sounds similar to Putnam’s statement about the relative unimportance of implementation properties for the understanding of mental functions: “*What is our intellectual form?* is the question, not what the matter is. And whatever our substance might be, soulstuff, or matter or Swiss cheese, is not going to place any interesting first order restrictions on the answer to this question” (Putnam 1975 302, italics in original).⁴⁹ However, Marr does not deny the relevance of ‘what the matter is’. On the contrary, for he continues his statement concerning the study of bird flight with the acknowledgement that here, too, the implementation level is relevant when other levels of analysis are attended to: “[i]n order to understand bird flight, we have to understand aerodynamics; only then do the structure of feathers and the different shapes of birds’ wings make sense” (Marr 1982 27).

In such a situation, knowledge about the implementation hardware can also provide us with clues about the nature of the algorithms that a device probably uses. For instance, since neurobiology shows that ‘wires’ or connections grow in brains rather quickly and cheaply and do so in three dimensions, parallel computing appears to be an attractive option, in contrast to the preference for serial computing by electronic devices and their two-dimensional printed circuit boards (Marr 1982 24). A confirmation of this preference for parallel computing that stems from such a fundamental principle of the biology of the brain has been the discovery of several streams of processing visual information in the human brain. This discovery is partly a result of neuroscientific studies and has had implications for the explanatory contributions that stem from other disciplines.⁵⁰ However, as was the case with the algorithmic level, where several options were found to be available for the same computation, the implementation level offers the investigator a similar room for choice. Not just because of the possibility that a particular brain or device may perform a certain computation in more than a single way, there is also a choice available in the granularity with which one looks at implementation. Focusing at a

⁴⁹ With the italicized question Putnam refers to the alleged Aristotelian preference for a formal explanation of the mind over a material one. He has in fact misjudged Aristotle’s hylomorphism, which in this context led Aristotle to be a precursor of embodied cognition theory (van der Eijk 1997).

⁵⁰ It appeared that distinct kinds of information are subserved by different streams of visual information processing (Goodale and Milner 1992), of which the number, their functions and their interaction remain a matter of debate (Creem 2001).

neural network level or at the finer-grained level of neurochemical interactions may give us different insights and accordingly different clues about the upper levels of the multi-level explanation of a function (Kosslyn and Maljkovic 1990).

As a result of the foregoing, we may conclude that explanations of cognitive functions involving the implementation level may be corrected without other aspects of the multi-level explanation automatically requiring correction as well.⁵¹ Obviously, this should not lead to the conclusion that the implementation level does not really matter in explanations of cognitive functions. On the other hand, drawing conclusions on the basis of insights that focus on the implementation level alone is always rash and should be treated with extreme caution. To close this section on Marr's methodology, we will therefore focus on his proposal for relating these three levels as part of a comprehensive explanation.

3.4 A loose interdependency between levels

Marr's methodology is primarily based on the conviction that for many cases of understanding complex systems: "one must be prepared to contemplate different kinds of explanation at different levels of description that are linked, at least in principle, into a cohesive whole, even if linking the levels in complete detail is impractical." (Marr 1982 20). As noted above, the levels involved in this account do not refer to entities in an ontological sense or levels of a mechanism – to be discussed in chapter I.5 -, but rather to different 'perspectives' (McClamrock 1991).⁵² The requirement to link such levels of description or understanding 'into a cohesive whole' leaves open the question how such linkage is to be carried out.⁵³

Marr dismisses the option of always requiring unique connections between the analyses, since at each level of analysis of a given function "there is a wide choice available" (Marr 1982 25). In explaining a certain function at several levels of analysis, researchers must therefore leave room for the possibility that an alternative computation or algorithm is carried out during the performance of a function, which may require a different implementation than previously thought. The benefit of such multiple realizability prevalent at all three levels of analysis is: "that since the three

⁵¹ Not surprisingly, however, especially Marr's ideas about object recognition have been criticized with respect to the computation and algorithmic levels and with respect to its implementation (Glennester 2007).

⁵² MacClamrock convincingly argues that Marr's levels have no ontological import but should be considered as three different perspectives on any given function or component function (McClamrock 1991).

⁵³ The following rendering of Marr's account leaves out this question of integration of the levels: "levels of analysis concern the conceptual division of a phenomenon in terms of different classes of questions that can be asked about it" (Churchland and Sejnowski 1988 741). The levels in Marr's approach are more than just a method for framing questions or a heuristic, as will be argued.

levels are only rather loosely related, some phenomena may be explained at only one or two of them” (Marr 1982 25). After all, the properties or constraints that appear at a particular level do not all transpire to the other levels of the function. For instance, although a computer may use completely different algorithms from a human brain, both systems may still perform the same task (Marr 1982 27). If we want to compare different systems, it is therefore very important to target the relevant level of analysis for such a comparison.

The loose interdependency also has as an advantage that the methods and results pertinent to all three levels do not always have to progress simultaneously. Equally, it makes each level relatively immune to flaws that are present at other levels, just like Marr’s dissatisfaction with the neurophysiological strategy of searching for feature detectors inspired him to put more efforts in developing an adequate computational or algorithm theory for the understanding of such neurophysiological findings. Understanding a particular algorithm can at times help to make progress in uncovering the associated mechanism, or vice versa. Their ‘loose independence’ does not exclude the presence of mutual constraints, where results at one level can guide research at another level, for example: “the form of a specific algorithm can impose strong constraints on the mechanisms, and conversely” (Marr and Poggio 1977 12) – the discussion in the previous section of biological wires and a preference for parallel computing confirms this observation.

Nonetheless, the role of mutual constraints between levels of explanation is less prominent than it is in the mechanistic explanation of a cognitive function. As we will see below, this is due to a different take on the idea of ‘levels’ itself. For Marr, the levels of explanation or analysis are perspectives on a particular function or component task of that function. Even though he emphasized the relative independence of these perspectives, it is important to note that one of the major contributions of his work has been to demonstrate the feasibility and necessity of integrating different perspectives: “Before Marr, researchers who studied artificial intelligence were concerned with ‘disembodied’ information processing, and paid scant, if any, attention to the brain; and researchers who studied the brain were concerned with neuroanatomy and neurophysiology, and paid little attention to formal analyses of what the circuitry does. These communities are no longer isolated from each other, and this ultimately may be Marr’s greatest achievement” (Kosslyn and Maljkovic 1990 250).

A similar integration of perspectives likely benefits the investigation and explanation of other cognitive functions, like singing or action determination. In those contexts as well, we may hope for a loose independency yet with some interdependency between the different levels of analysis – or disciplinary perspectives

- involved. To make our hopes more robust, however, it is useful to end this chapter with a short evaluation of Marr's approach.

3.5 Modularity and some limitations of Marr's methodology

Even though Marr strongly defended his approach, he was quite aware of some of its limitations. He avoided taking an extreme position, for example, with respect to the predominance of one level of analysis over the others. Having distinguished different levels of analysis and associated these with different strands of research, he demonstrated how a careful navigation between these levels bolsters our understanding, whereas partisanship for a particular level risks unnecessarily weakening it.⁵⁴ Instead, the combination of these levels is what furthers comprehensive understanding of a particular function or function component. What such a combination seems to require, however, is the relative isolation of such a function or function component. Only on the basis of a certain 'modularity' of a given object of investigation can researchers apply their multi-level analysis to it, Marr assumed. This he thought to be a prime reason for the limitations of his approach. We will elaborate on this assumption of modularity and at the end of this section link it again to his advocacy of loose independence between levels of analysis.

As noted earlier, Marr's main research was on early vision, a stage of visual information processing which he believed to be relatively insensitive to other functions of the organism or to later stages of visual perception.⁵⁵ This assumption met with criticism on both theoretical and evidential grounds. Marr was not naïve in this matter and realized that the assumption might turn out to be inappropriate. Supporting the assumption, though, was the consideration that the evolution of a complex function would generally benefit from avoiding the situation in which it was vulnerable to many small and local changes in the overall system.⁵⁶ Correspondingly, he assumed that the evolutionary development of such a function could better not interfere with earlier developed component functions.⁵⁷ Obviously, later evolved

⁵⁴ Marr's emphasis on the computational theory has nonetheless made some authors argue that he denied the implementation level's relevance (cf. Glennerster 2002 ; Polger 2004 ; Sun, Coward et al. 2005). We think that (Kosslyn and Maljkovic 1990) are more correct in stating that Marr's works demonstrate an imbalance that he theoretically did not justify.

⁵⁵ This assumption of early vision's peculiar nature was disproved many years ago. See e.g. the more recent account of V1's function and place in the architecture of vision in (Ng, Bharath et al. 2007) which includes not just feed-forward but also several feed-back relations to V1.

⁵⁶ This estimation of interaction may be due to Marr's idea that the stages make up a sequence-like process, where information must always pass through these stages (Glennerster 2002).

⁵⁷ The correspondence of evolution with modularity is indeed plausible, as is argued in (Wimsatt and Schank 2004), although they emphasize the importance of distinguishing between different forms of modularity in this context.

functions could take up tasks in addition to tasks that had already been performed by functions that evolved earlier. Indeed, he believed that: “as evolution progressed, new modules came into existence that could cope with yet more aspects of the data” (Marr 1977b 41).⁵⁸ To capture this idea of developing functions more generally, he formulated a ‘principle of modular design’ (Marr 1976 485; Marr 1982 102) which states that breaking up a large computation into small and independently executed ones makes it more robust.⁵⁹ In addition to its robustness, such a structure would also enhance the speed and fluency of the process,⁶⁰ which could be hampered by hierarchical interactions (Marr 1976 502).⁶¹ In the second part of this dissertation, we will discuss empirical hypotheses regarding the prevalence of modular structure in a cognitive system like the brain as a result of its evolution and ontogenetic development. For it has been argued from various perspectives – computational, evolutionary, developmental, to name a few – that such a structure yields several benefits to the organism or system that is equipped with such a structure.

However, apart from such benefits, Marr also had some epistemological considerations, believing that a truly complex system with many interactions between different stages of its processing trajectory might be impossible to understand.⁶² Indeed, in the ‘Conversation’ added to his book on vision, he even admits that his approach would fail with: “Systems that are not modular. Things like the process by which a chain of amino acids folds to form a protein - that is to say complex, interactive systems with many influences that cannot be neglected.”

⁵⁸ As we will discuss more below, the ‘Massive Redeployment Hypothesis’ argues that evolution plausibly also proceeds by re-using components that are already in place: evolution proceeds by redeploying and adjusting older components of an organism, suggesting integration rather than mere modularity as the *modus operandi* of many components (Anderson 2007).

⁵⁹ Marr did not articulate the notion of modularity as Fodor did around that time. Still, there are clear affinities between both notions. Indeed, several of Fodor’s defining properties of modularity would apply to Marr’s theory of early vision. The Fodorian characteristics of a modular system that seem to apply to early vision are its being a mandatory input system, its providing ‘shallow’ outputs to later stages of vision, and its being a system that is hardly accessible by other cognitive functions and working fast. Other Fodorian properties of modularity, like the informational capsulation of the coupled visual-motor system or the association with a fixed neural architecture and with specific failures do not have direct parallels in Marr’s work or are not discussed there (Fodor 1983).

⁶⁰ When discussing language, Marr expected it to be modular in structure because of the characteristics of its ‘fluency’, its smooth continuation and the absence of conscious attendance or interference (Marr 1982 356). However, some hesitation is aired in that context: “It’s not clear, and some claim it’s inherently not modular and should be viewed much more heterarchically” (Marr 1982 356).

⁶¹ As we know, quite the contrary –that sensori-motor coupling makes vision an easier task- has been extensively argued in i.a. (Clark 1997 ; Noë 2004 ; Thompson 1995). Marr may have underestimated the benefits that a hierarchical structure yields to a complex and dynamic system. Or, to be more specific, the benefits that a heterarchical structure yield – see footnote 96 on heterarchy.

⁶² Simon confesses that he is unsure about the chicken or the egg regarding his epistemological preference for a particular structure of a complex system and his argument that such a structure must have been more profitable for the system during its evolution (Simon 1962).

(Marr 1982 356).⁶³ Similarly, he adds, syntax appears to be almost modular although it may have sparse interaction with semantics, necessary to narrow down syntactical information in some cases. From that short discussion we can derive that Marr fears that a system or function which is not part of a strictly hierarchically structured process may be very hard to analyse and explain. Such a function, whose processes would then be partly determined by continuous interactions with other functions, is hard to study and then explain in isolation from those other functions. Due to its complex structure, a function could turn out to be modulated by other functions or component functions at stages that were originally assumed to be irrelevant to it. We may be forced to expand the function under study and include ever more functions in it that were previously thought to be distinct from it – making it ever harder to analyze and explain it.

The result of these considerations regarding a function's consisting of isolable modules corresponds in a certain sense to the earlier considerations about the relations between the levels of analysis in Marr's approach. As much as the computational, algorithmic and implementation levels of analysis of a function were found to potentially constrain each other's results and guide each other's investigations, they would still allow researchers to offer separate explanations. Only because of this separation of different levels of analysis or different perspectives on the function under study, would flaws or misunderstanding at one level not automatically affect the results at other levels. We may conclude that in Marr's view, the robustness of a system or function with a modular structure corresponds with the robustness of the scientific results of its study when these results do not rely too much on each other.⁶⁴ Given the situation in cognitive neuroscience nowadays, in which functions are allegedly subserved by heterarchically structured neural networks and in which notions like embodied or enactive cognition reign, one may well wonder why Marr's approach still receives the attention and support it does. Indeed, when Ochsner &

⁶³ Elsewhere Marr discussed the distinction between two types of theories that are employed for solving problems in artificial intelligence. Type-2 solutions describe complex and interactive problem solving processes. At the time (in 1977), Marr advised not to concentrate on such difficult problems, while noting that many authors focused on problems that humans perform poorly even though we may readily understand these problems intellectually. For instance, they focused on arithmetic, even though Marr does not believe that we mentally perform along the lines that arithmetic problem solving is described. Ironically, he concludes that "one is left in the end with unlikely looking mechanisms whose only recommendation is that they cannot do something we cannot do" (Marr 1977b 45).

⁶⁴ It has been argued that multi-level systems can also be analyzed and explained in terms of a hierarchical model structure, where a model that accounts for a sub-system at one level can constrain the number of plausible models at another level. The authors explicitly parallel their hierarchical model structure with Marr's multi-level approach (Meeter, Jehee et al. 2007), although in our view the comparison denies the quite distinct nature of Marr's implementation level from the other two levels.

Kosslyn list their ‘five general points about cognitive functions’ as they are studied in cognitive neuroscience, it appears as if their list contains several objections to Marr’s approach. This holds particularly for those referring to the highly interactive nature of the neural networks that make up cognitive functions, which are implemented in distinct brain areas and are processing information not just serially but also in parallel (Ochsner and Kosslyn 1999 354). Similarly, it has been argued that both the phenomenology of perception and the apparent dynamical processes that underlie perception (Borrett, Kelly et al. 2000) are at odds with the modularity assumption, which we found to be prominent in Marr’s approach. The development of parallel distributed processing and the discovery of such processes in the brain’s networks have also suggested that some of the advantages that Marr attributed to modularized systems – like speed, robustness, and evolvability – can actually be exhibited by those neural networks.⁶⁵ However, the suggested opposition between the assumption of modularity and the parallel distributed nature of neural networks itself deserves to be considered along the lines of the approach Marr defended.

According to that approach, we should not be surprised to find that a computational theory of a given function – and its psychophysical or cognitive psychological investigation – involves a certain degree of modularity of that computation or the algorithm that performs the computation. Nonetheless, the implementation theory of the relevant computation or algorithm must be assumed to be only loosely related to theories regarding those other levels, leaving ample room for implementation by a network structure that shows a more limited form of modularity. Assuming otherwise, that an implementation by a parallel network would contradict a modular structure of a function, is usually based upon the unwarranted conflation of the anatomical structure subserving a function with its functional structure as described by its computational theory.

Even though the assumption that cognitive functions are subserved by the parallel distributed networks that are prominent in the brain has gained ever more prominence, the distinction between functions and components functions has survived. More and more, the results of evolution, development and learning are being described in terms of the acquisition of specific modules or the development of a function’s modularity. In Part II we will see how and why this is the case, when we consider

⁶⁵ Paul Churchland has argued extensively about the characteristics of such parallel distributed networks or connectionist approaches for a variety of cognitive functions in his (Churchland 1995), while Patricia Churchland many years earlier paved that way in her seminal (Churchland 1986) in which she notes that the parallel distributed networks approach grew partly because of the unsatisfying results with serial computational approaches. Marr’s approach was more of the latter fashion, indeed. Had Marr lived longer, he probably would have agreed with that.

why it is plausible for dynamic and evolving systems to develop intermediate stable forms (Simon 1962), generative entrenchments (Wimsatt 1986), or kludges (Clark 1987). We will explore empirical cognitive scientific theories about ‘modularization’ of cognitive functions that takes place during child development (Karmiloff-Smith 1992) and the formation of distinct processes or functional systems – ‘kludges’ - for extensively practised cognitive functions.

In sum, even though several of Marr’s specific theories concerning – components of - vision and some of his methodological assumptions do not find adherents anymore, some of the general principles of his approach are still considered relevant and beneficial to a ‘flourishing’ neuroscience (Rolls 2011). As we turn to the next methodologies to be discussed in this part, the reader may discern how Marr’s approach has been integrated in other and more recent ones. Moreover, since these approaches are being applied to cognitive functions that are considered to be ‘higher’ and more complex than the early and middle stages of visual information processing to which Marr limited his research, their discussion prepares us for the investigation of the domain of this dissertation, concerning the determination of human action.