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

Keestra, M.

Publication date
2014

5 MECHANISTIC EXPLANATION AND THE INTEGRATION OF INSIGHTS*

The three previous methodologies were found to differ in several ways, a major difference being the role assigned to conceptual, empirical or other – algorithmic, for example – insights in the explanation of a cognitive phenomenon. A discussion of these methodologies has among other things yielded the result that pluralism in such an endeavor is inevitable. We noted that defining a complex phenomenon like singing or consciousness will already confront the researcher with likely pluralism. Devising an algorithmic theory – or computational model - for a particular cognitive task similarly offers a plurality of options. Finally, a causal plurality is general involved in a complex phenomenon, to which consequently several distinct theories equally apply, although each with only a limited explanatory power.

A result of this discussion is the need for a methodology that can handle such pluralities, that allows for a phenomenon being given divergent or incomplete definitions, that can integrate different types of theories, and that can handle causal pluralism. The aim of the next and last methodology to be discussed is to present mechanistic explanation as a useful approach that fulfills these desires. Mechanistic explanation requires researchers to determine how their specific explanatory or theoretical insight fits into a so-called 'explanatory mechanism' of a particular phenomenon. A definition for such an explanatory mechanism states that it is: "a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena" (Bechtel 2008 13). Fitting an insight into an explanatory mechanism implies first scrutinizing whether it applies to a particular component part or operation, or to an organization feature of the mechanism. Further below we will shed light on how this might work.

The merit of this mechanistic explanatory approach is in our eyes its ecumenical yet not undemanding nature, supporting what Craver calls: "the mosaic unity of neuroscience" (Craver 2007). Notwithstanding its liberal stance with regard to several forms of pluralism, it does require researchers to determine how their specific insights are to be integrated with other available insights in a phenomenon. Another merit is that this approach does not suggest having unlimited applicability. On the contrary, the definition of an explanatory mechanism immediately provides a first

* On pages 371, 373, 375 figures I, II, III offer simplified representations related to the arguments made in parts I, II, III respectively. Figure I is particularly relevant as a represention of the main contents of this chapter I.5.

caveat: it is relevant for the explanation of one or more specific phenomena, nothing more and nothing less. We will meet a second caveat later: this methodology does not make the metaphysical or epistemological claim that all phenomena allow such a mechanistic explanation. In fact, there is good reason to assume that there are components in each phenomenon – if only the absolutely smallest ones – that are not explainable in this way.[86] Nonetheless, for the present context and for the explanation of how an agent is capable of 'sculpting the space of actions', this approach offers satisfactory means. In order to appreciate these resources, let us recapitulate some relevant insights from the previous sections.

The last sections devoted to the NCC approach highlighted its liberal use of strategies and heuristics in the investigation and explanation of a phenomenon as elusive as consciousness. Nonetheless, the mere finding of a neural correlate was said only to make sense when it can be interpreted as an ingredient of a more comprehensive mechanism (Chalmers 2007). Other heuristics have been employed by the other approaches. Marr, for example, subscribed to a 'principle of modular design' which implies that "a large computation can be split up and implemented as a collection of parts that are as nearly independent of one another as the overall task allows" (cf. Marr 1976 485; Marr 1982 102). Applying such a principle does not stand in the way of recognizing its limitations, for instance that it "does not forbid weak interactions between different modules in a task, but it does insist that the overall organization must, to a first approximation, be modular" (Marr 1982 102). In the present, mechanistic explanatory, approach, related principles or assumptions are made concerning the structure of a cognitive process and its explanation, as we will see below.

Another strategy that Marr emphasized was distinguishing between and subsequent integrating 'three lines of attack' (Marr 1982 300) for the analysis and explanation of a cognitive task. In combining three different perspectives – a task analysis, algorithmic theory and neural implementation theory – he invited the interdisciplinary *integration* of several disciplinary perspectives on such a task while allowing the distinct endeavours also some *independence* (Marr 1982).

A third strategy that we discussed above prescribed to first analyze and define the cognitive function or computation under scrutiny. Setting aside their criticism of Marr, Bennett & Hacker agree with him on the importance of such a definition (Bennett and Hacker 2003).[87] As ingredients of such a definitory effort, they specifically mention the use of conceptual analysis and the behavioral criteria that

[86] See the discussion of limitations of mechanistic explanation in section I.5.7.

in their view guide our use of the concept of a psychological function. In addition to their view, we argued in section I.2.3 that conceptual variabilities or blurred distinctions between functions (as in blindsight or inattention to pain) can also be used as heuristics, as they sometimes refer to phenomena that are determined by interferences between two different functions or other modulations of a particular function. Our example of singing helped us to spell out such variabilities, underlining the difficulty of providing a single definition of such a function. A requirement of an explanation of such a function is that it helps to provide insight in these variabilities or modifications, which is something that a mechanistic explanation is capable of.

In addition to the lessons drawn from previous approaches, another element needs to be added to the ingredients that a comprehensive explanatory account should be able to integrate. If it is to analyze and explain how development and learning take place, it should be able to accommodate environmental influences, ranging from sensory stimuli to resources that stem from other agents, teachers, etcetera. In contrast to the closed 'behavior program' of the parasitic cowbird, for example, young geese display an open behavior program, as they would continue to associate with humans after being raised by Konrad Lorenz instead of a goose: their imprinting mechanism is to a large extent open for environmental information (Mayr 1964). Indeed, it has been argued that in all dynamic and evolving systems the distinction between innate and acquired traits should be considered gradual and not disjunctive, as environmental information will always become effectively integrated in those systems (Wimsatt 1986). As we will elucidate in Part III how an agent's 'space of actions' can also become constrained by higher order intentions – sometimes stemming from joint deliberation with another agent - we are currently interested in an explanatory approach that permits such influences to determine an explanatory mechanism's behavior.

These requirements and the employment of the strategies that previous approaches have yielded are not easy to fulfil by any account. However, the requirements may gain in plausibility if we illustrate such explanatory efforts with reference to skill learning. Obviously, motor or cognitive skill learning involves many different

[87] Their rejection of applying the notion of 'representation' to anything other than the symbols that humans use to represent things and thus of applying this notion to brain functions (as Marr does) unnecessarily constrains its use and overlooks its *heuristic* use in the study of such brain functions (cf. Bennett and Hacker 2003 70, 76). For recent arguments for and against the use of the concept of representation as part of functional explanations in cognitive neuroscience, see for example (Bechtel 1998 ; Churchland 2002 ; Haselager, de Groot et al. 2003 ; Jacobson 2003 ; Keijzer 2002 ; Piccinini 2008). We will not discuss this issue explicitly, but in Ch. 3 we will argue that verbal representations play a functional role in the mechanism responsible for action determination.

functions, and their complex interactions are modified via dynamic processes like instruction, learning and experience. As a result, automatization of a motor or cognitive skill corresponds to changes in terms of the intentions, consciousness and control involved in it, or in terms of its goal and stimulus dependency, or in terms of its efficiency and speed (Moors and De Houwer 2006). Clearly, an explanation of such a process will be complex and will display intricate dynamic interactions between its many psychological or cognitive ingredients. Similarly, an explanation that focuses less on psychological but more on neuroscientific ingredients will refer to a gradual process involving a complex interaction between cortical and sub-cortical networks during learning and during automatized responses (Ashby, Turner et al. 2010).

Interestingly, such dynamical processes often lead to results that comply with Marr's 'principle of modular design', which was mentioned above. However, what Marr may not have realized is that in many cases modularity is not the starting point but rather the result of a dynamic process, for which the term 'modularization' has even been coined (Karmiloff-Smith 1992).[88] In fact, the mechanistic explanatory approach is particularly well-equipped to yield insight in such a dynamic process, as we will discuss in section I.5.6 below. This fact may to some extent fight Marr's doubt whether dynamic systems allow explanation, believing as he did that his own methodology could not apply to non-modular systems (Marr 1982). Indeed, he held that the Type-2 theories that could explain such complex processes were not available at the time (Marr 1977b). Since then, several such approaches have been developed, especially to cope with such processes, among which the mechanistic explanatory approach which has gained ever more recognition in cognitive neuroscience.[89] If,

---

[88] In his afterword to the new edition of Marr's posthumously published book on vision, his former collaborator Poggio speculates that Marr would have wanted to add to it the process of learning – having been the object of Marr's earlier research – if he had had time (Poggio 2010). One may also speculate how such an expansion of his approach would have required an extension of the methodology, as it is not easily applicable to the complex and interactive systems that demonstrate learning.

[89] Other explanatory approaches that need to be mentioned in this context are those that use Bayesian or probabilistic analysis of processes - like vision, for example (Yuille and Kersten 2006) – or those that commend the use of dynamic systems theory (van Gelder 1998). It can be argued that an explanatory mechanism to a large extent is similar to a dynamic system, depending on whether one assumes that representations play a role in such systems – as was defended by Bechtel (Bechtel 1998) and further refined in (Nielsen 2010). However, dynamic systems theory usually suffices with describing the behavior of a system, not explaining it, which is what a mechanistic explanation has to offer in addition (Kaplan and Bechtel 2011). The role of representation in explanations is by no means a settled issue, nor is the role of representations generally in cognitive neuroscience. First, it should be recognized that representations play different roles (Keijzer 2002). Similarly, the difference between perceptual and conceptual representations deserves to be acknowledged as a way to accommodate critique of representationalism (Markman and Dietrich 2000). An account that seeks a middle road between representationalism and non-representationalism has been suggested, among others, by (Gärdenfors 2004b) who proposes to replace representations with multi-dimensional conceptual spaces – bearing some similarity with the 'space of actions' discussed in this dissertation. (Haselager, de Groot et al. 2003) on the other hand argue generally that the ill-defined notion of representation renders this debate without much use.

moreover, dynamic processes lead to increasing modularization, Marr's hesitations do not need to withhold us from trying to explain such processes.

Let us therefore now discuss how the mechanistic explanation proceeds, integrating the ingredients mentioned in the present section. We will do so by highlighting its three heuristics or steps: definition, decomposition and localization. But first we will give a short introduction of the methodology of mechanistic explanation.

5.1 From the mechanicist world picture to the method of mechanistic explanation

Mechanistic explanation aims to explain a phenomenon by providing insight into a mechanism that is responsible for producing it. Whether it is a particular cognitive or behavioral phenomenon, a particular gen expression, or even the warming up of our atmosphere, many phenomena allow mechanistic explanation, thus providing insight into component parts and operations of an organized structure that together produce such a phenomenon (Bechtel 2008). Or, in slightly different terms: "[t]his is a mechanism in the sense that it is a set of entities and activities organized such that they exhibit the phenomenon to be explained" (Craver 2007 5).[90] Below, we will further explain how scientists gather and organize insights into such a mechanism that helps them to explain a particular phenomenon.

The clause that a mechanistic explanation pertains to a particular phenomenon is not unimportant. This specific mechanistic explanatory methodology should be distinguished from the long tradition in the history of science that considers nature as a whole as a mechanism. Indeed, 'mechanicist' world pictures and explanations have been en vogue from antiquity[91] up to modern scientists like Copernicus,

[90] There are still other definitions available of the eventual result of a mechanistic explanation – of an explanatory mechanism, that is. As the essential ingredients – a mechanism being constituted by explanatory relevant component parts and operations and their organization – are common to most influential definitions we will not go into the differences (see also definitions in Glennan 2002 ; Machamer, Darden et al. 2000 ; Woodward 2002). While establishing a comparative table of definitions – including some different from those noted here – Hedström comes to a similar conclusion: "Underlying them all is an emphasis on making intelligible the regularities being observed by specifying in detail how they were brought about" (Hedström 2008 321). In another review, he summarizes the telos of mechanistic explanation as: "proper explanations should detail the cogs and wheels of the causal process through which the outcome to be explained was brought about" (Hedström and Ylikoski 2010 50.

[91] An astonishing yet still relatively unknown example of an ancient mechanism is the portable Anthikythera-mechanism, dating back to the second century BC and having no equal within the next millennium. Found by sponge divers around 1900 and first described and partly explained only decades later (Price 1974), the mechanism still surprises researchers, who keep on discovering more astronomical and calendar calculations that it can perform, cf. (Freeth, Bitsakis et al. 2006). We would contend that it also offers a demonstration of the discontinuity of scientific progress, with antique scientists like Archimedes and Aristotle having no immediate descendants until the scientific revolution some two millennia later.

Huygens, Descartes, Laplace, Boyle and Newton and beyond.[92] Inspired by artificial machines or mechanisms, these traditional mechanicist authors tried to reduce all complex phenomena that reality presents to us - usually exclusively material reality, that is – to the products of many, yet simple parts and movements. Most authors recognized that not all parts in nature would allow such a reduction, as it would lead to an infinite regress. However, this did not stop the plea to eventually 'dissolve' all physics in mechanics, as voiced in the slogan that: "all physical appearances have to be explained with reference to natural forces that are exerted by material points upon each other" (Dijksterhuis 1969, translated from Dutch original, p. 538). This metaphysical conviction is specific to this traditional mechanicist world picture.

That is not to deny that there is a methodological interest that contemporary mechanistic explanations share with the traditional mechanicists. Allowing for the importance of the discovery of laws of nature and the probable causal connections behind those laws, both aim to elucidate a mechanism that might be responsible for particular phenomena.[93] However, a crucial difference between the current view on mechanistic explanation and the traditional mechanicist world picture concerns the metaphysical claims that accompanied the latter.[94] Unlike these traditional mechanicist explanations, scientists who currently develop mechanistic explanations do not subscribe to a reductionist and atomist agenda.[95] On the contrary, they do not focus on such metaphysical issues and even "reject any fixed and limited list of the modes by which parts of mechanisms can act and interact" (Glennan 2008 377). This has to do with the recognition that complex mechanisms often have some properties that cannot be explained with reference to such a fixed – and limited - list of parts

[92] Obviously, there are important differences between these authors that cannot be articulated in the scope of this dissertation. For example, the implications of Descartes' distinction between extended matter and 'res cogitans' could be relevant. Similarly, Newton's acceptance of gravity as an essential force without an apparent underlying mechanism may be due to his religious perspective, trumping his 'mechanicist' convictions. As for gravity, perhaps a graviton allows its mechanistic explanation, but perhaps the approach is not applicable to this important phenomenon. That would not exclude its applicability to most other phenomena in the material world, though.

[93] Authors defending mechanistic explanation usually contrast this approach with modern deductive-nomological explanation, which implies that an explanation of a particular phenomenon is deducible from at least one natural law. Corresponding with this contrast is the authors' interest in phenomena that do not behave completely law-like and exceptionless (Bechtel and Abrahamsen 2005 ; Craver 2007 ; Machamer, Darden et al. 2000). However, Leuridan warns against focusing solely on explanatory mechanisms while neglecting the observation of regular or lawlike phenomena, since in the life and cognitive sciences such regular phenomena can be multiply realized (Leuridan 2010).

[94] It is therefore not the case that each and every discovery of a causal nexus requires a mechanistic explanation before it can be recognized as such. Indeed, not all discovered causal connections will eventually allow such an explanation, in some cases the 'black box' will remain closed. Besides, it is advisable, for example, to treat an epidemic for which a causal factor has been found, even if no explanatory mechanism has been discovered (Broadbent 2011).

and interactions.

To understand this, it is useful to acknowledge the difference between aggregate, composed and evolving systems. An aggregate system is constituted by a mere collection of parts with simple interactions, such that the addition, subtraction or substitution of parts will not have any effect on the properties of that system. As Wimsatt has argued, this condition holds only for a system's mass, while even the volume of combined volumes of sugar and water for example is not merely aggregative. A system's non-aggregativity is most visible in its having emergent properties (Wimsatt 2007).

In regard to most of their properties, most systems are indeed not just aggregative but composed instead. This implies that to explain most of their properties, scientists must also take into account the organization and interactions that occur between a system's component parts and operations. It is due to such organization and interactions that a composed system displays emergent properties that are irreducible to its smallest components (Bechtel and Richardson 1993). Moreover, a complex system generally displays a hierarchical form of organization in which levels can be distinguished, as this yields advantages in terms of the system's stability and response speed, among others (Simon 1962 ; Wimsatt 2007).[96] This organization structure not only has implications for its interactions with its environment generally, but also for its scientific investigation. Since investigations often involve intervening with a system's properties and detecting the consequences, it will have to take into account the differences between a system's levels and their corresponding properties.

Environmental interactions are not equally important for all systems. Obviously, a sugar solution has a more differentiated response pattern regarding environmental

---

[95] Not all authors subscribing to mechanistic explanation equally shy away from statements with ontological implications. Most explicit in this respect is Wimsatt, who writes about organizational levels in nature, which are occupied by: "families of entities usually of comparable size and dynamical properties" and to which explanations generally refer (Wimsatt 2007 204). Critical of such observations is Craver who denies that levels of an explanatory mechanism are identical to such levels of nature or that the size of the respective entities matters at all. His central point is "that levels of mechanisms are defined componentially within a hierarchically organized mechanism, not by objective kinds identifiable independently of their organization in a mechanism" (Craver 2007 191). Craver's emphasis on this point makes him perhaps oblivious to the intriguing fact that there is a parallel – though not an absolute – between these levels of nature and levels of mechanism, to which Wimsatt draws attention.

[96] Or, to be more precise: heterarchical – a term introduced to refer to a structure of neural networks (McCulloch 1945). In heterarchically organized systems it can occur – due to learning, for instance – that in an explanatory mechanism a particular component at an intermediate level may be bypassed during the performance of a function, compared to the mechanism before such modification (Berntson and Cacioppo 2008). Such heterarchy is observable not just in organisms or the brain, but also in societies, for example, which display more variability and change in power relations than hierarchical structures would allow (Crumley 1995). When reference is made in this dissertation to hierarchically structured and dynamic systems, it is implied that these are in fact heterarchical structures.

changes in temperature, pressure, chemistry, than either sugar or water alone. However, this environmental interactivity or context sensitivity is exponentially greater in systems that can evolve or develop than in aggregate or composed systems.[97]As we will see in this dissertation, the composition of an evolving and developing system is such that it is modifiable as a result of interactions with its environment. In section I.5.6 below, we will emphasize how mechanistic explanation is particularly suitable for the explanation of changes in cognition or behavior due to development and learning, as it can refer to modifications of the explanatory mechanisms. It would be hard or impossible to account for such changes with the ingredients of the traditional mechanicist world picture. For this reason, it offers very limited resources for the analysis and explanation of the regular yet not exceptionless phenomena that pervade the life and social cognitive sciences (Glennan 2008).[98]

Given their suitability for explaining complex and dynamic phenomena, investigations in these fields have often led to the development of explanatory mechanisms, even if not always explicitly so. In their seminal and extensive exposition of the approach, Bechtel & Richardson offer many examples of such research histories: from the Krebs cycle to genomic regulation, from fermentation to cognitive psychology, scientific investigations have resulted in mechanistic explanations while employing the specific heuristics of decomposition and localization (Bechtel and Richardson 1993). More specifically to cognitive neuroscience, similar analyses have been presented for the mechanistic explanation memory (Craver 2002 ; Craver and Darden 2001), for vision (Bechtel 2001b), for action understanding (Keestra 2011 ; Looren de Jong and Schouten 2007), for circadian rhythms (Kaplan and Bechtel 2011), for example. More recently still, mechanistic explanation is being considered as a methodology for the social sciences, with examples referring to social phenomena like the self-fulfilling prophecy or network diffusion (Hedström and Swedberg 1996 ; Hedström and Ylikoski 2010 ; Tilly 2001).

---

[97] The difference in notions of emergence can be generally ascribed to the different – external or internal – contexts with which a system's interactions give rise to emerging new properties (Wimsatt 2006a). For our present purposes, this differentiation is not relevant.

[98] Beatty argues why the converse model – a singular theoretical account of a biological phenomenon – is unlikely and why in a theoretical pluralistic model of explanation there are differences in relative significance between theories (Beatty 1997). Both arguments concur with the mechanistic explanatory approach considered here. 87 With regard to the context of social sciences, it is especially a matter of dispute whether entities such as the state, religion, or collective memory can enter as component parts into an explanatory mechanism, or that these have to be considered only as environmental factors of such a mechanism. Similar questions arise as to the wide variety of social interactions that social scientists include in their explanations of social phenomena: are power or sexual relations fit to be integrated as component operations in a mechanistic explanation of social phenomena? Hedström & Ylikoski, for example, argue for the development of mechanistic explanations in the social sciences with a crucial role for individual agents and their relations (Hedström and Ylikoski 2010).

For each of these subjects, the question as to what component parts and operations are involved in the responsible explanatory mechanism is of primary importance.[99] Moreover, given the non-aggregative nature of such a mechanism, it is likely constituted by parts and operations organized at different levels. Before we move on to a systematic treatment of the methodology of mechanistic explanation via its three heuristics, we will briefly discuss the explanation of memory as an example of this particular method.

5.2 Memory and the mechanistic explanation of learning

Since memory and learning are essential phenomena in cognitive science and are also relevant for this dissertation, with its interest in the process of an agent's 'sculpting' the space of his actions, let us consider these as examples for the endeavour of mechanistic explanation. To explain learning accordingly, researchers should provide an ever more comprehensive description of the mechanism that is responsible for it, consisting of component parts and operations and their organized functioning (Bechtel 2008). We must always avoid the assumption that by presenting an explanatory mechanism we are simultaneously providing an exhaustive and exclusive description of all possible functions of the components involved. For especially in complex and dynamic systems that have both an evolutionary and developmental history, it is usually the case that a component is involved in more than just a single function, like it is the cases with a gene that can be co-responsible for several phenotypical properties. Indeed, it has been argued that 're-use' of neural components is a prevalent phenomenon with regard to the brain, implying that many parts and operations figure in more than a single cognitive functions (Anderson 2010).[100]

In the case of memory, cognitive scientists have devoted many investigations to

[99] With regard to the context of social sciences, it is especially a matter of dispute whether entities such as the state, religion, or collective memory can enter as component parts into an explanatory mechanism, or that these have to be considered only as environmental factors of such a mechanism. Similar questions arise as to the wide variety of social interactions that social scientists include in their explanations of social phenomena: are power or sexual relations fit to be integrated as component operations in a mechanistic explanation of social phenomena? Hedström & Ylikoski, for example, argue for the development of mechanistic explanations in the social sciences with a crucial role for individual agents and their relations (Hedström and Ylikoski 2010).

[100] Relatively independently and based upon different lines of evidence, several hypotheses have been presented in recent years that are comparable to Anderson's hypothesis regarding extensive neural re-use in the brain (Anderson 2010). For example and building upon evidence about mirror neurons, it has been argued that neurofunctional architecture is being 'exploited' for more than just a single function (Gallese 2008), while cognitive neuroscientific research of reading and writing – which are relatively recent cultural inventions – has suggested that evolutionary older neural circuits are being 'recycled' (Dehaene 2005).

spatial memory in mice. Craver has argued how these studies have culminated in a mechanistic explanation of this phenomenon – and the present section is based upon his presentation in (Craver 2002 ; Craver 2007 ; Craver and Darden 2001). Studying spatial memory, researchers let the animals wander through labyrinths or in the Morris water maze, in which a platform is hidden in opaque water, and measure the time and trajectory they use for finding their way to this platform. Apart from such behavioral measures, researchers can also measure the activities in different brain areas of the animals and try to disentangle areas specifically involved in spatial memory and not responsible for motor behavior or visual processing, for example. Or they can zoom in on a particular area and detect the electrophysiological interactions that occur in the synapses within that area and try to correlate that with particular phases of the animal's behavior. Digging even a level deeper, researchers have investigated the molecular processes that constitute the synaptic electrophysiological

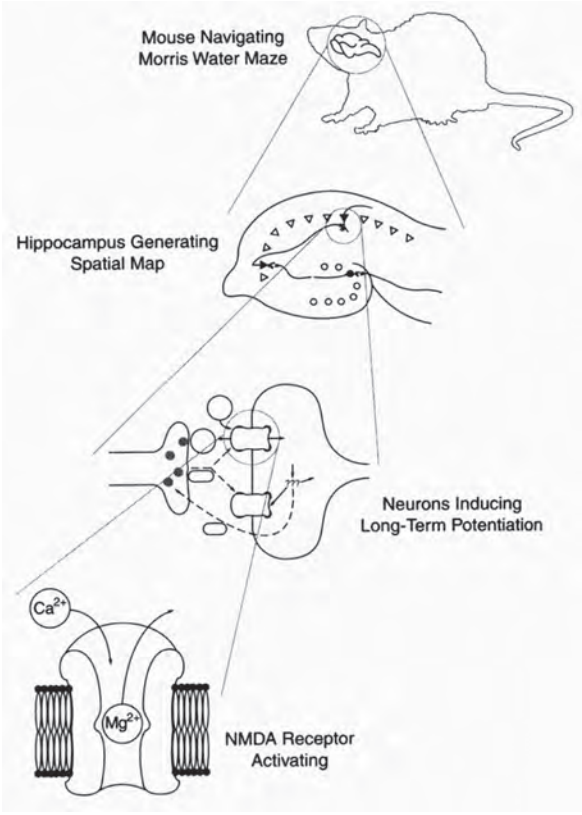

Figure 1: levels of spatial memory. Reprinted from (Craver 2007 166) with permission from the publisher

interactions corresponding with the process of Long-Term Potentiation (LTP). This LTP, in turn, occurring in the relevant brain areas under specific environmental conditions, is responsible for the animal's spatial memory learning and the increase of successful navigation. Figure 1 presents a crucial component of the relevant explanatory mechanism of memory, represented at different mechanism levels.

Though we cannot describe the contents of this figure in detail, it may help us to explain what is at stake in mechanistic explanation. Figure 1 comprises figures at four different levels, which all refer to the same phenomenon, namely spatial memory. Spatial memory as such, however, can only be observed as a phenomenon in the mouse's navigating efforts, represented at the top level. Simply put, the main mechanism that was found to be responsible for this memory was located in the animal's hippocampus. So it was this part of the animal rather than other neural or even muscular parts that turned out to be specifically responsible for spatial memory. Obviously, this component part cannot produce successful navigation on its own, as the hippocampal cells still need to interact with neural areas responsible for motor planning, visual perception, and so on. These interactions consist of excitations and inhibitions between neural areas, that together produce the navigation behavior. Research on mice with lesions in the hippocampus has shown that they have specific difficulties in learning to navigate, suggesting a specific function of this neural area.

Knowing that the hippocampus has a specific function for spatial memory does not yet provide insight in the processes that occur in that area. Nor does it shed light on the specific temporal patterns of mice's learning or the sensitivity of this learning process for specific chemical interactions. As Craver argues, the discovery of the sub-mechanism of Long-Term Potentiation as a general synaptic process constituting learning also made it possible to explain specific properties of learning. For example, it turned out that the response – i.e. depolarization - properties of post-synaptic cells correlate with the difference in learning results when an animal is exposed to rapid and repeated stimulation in comparison to single exercises ((cf. figure 5.2.c and text in Craver 2007 168-169). Even though reference to hippocampal activities and their interactions with other areas would be largely sufficient to account for the capability of spatial memory, its temporal peculiarities were therefore better explained by referring to the interactions between singular cells and their changing properties.

These cell properties and interactions are in turn constituted by molecular processes. For the intervention in spatial memory, we can target these processes, if irreversible surgical lesions or difficult electrophysiological lesions have to be avoided. To that effect, we can block the NMDA receptor's channel with particular

chemicals, and thus interfere with the LTP process that is required for successful spatial memory. Here, the component parts are not neural areas or cells, nor are they operations like excitation, inhibition, polarization. At this molecular level we find instead "entities like the NMDA and AMPA receptors, glutamate, Ca2+ ions, and Mg2+ ions engage in activities like attracting and repelling, binding and breaking, phosphorylating and hydrolizing" (Craver 2002 S89-S90).

As a result of mostly separate ethological, neuroanatomical, physiological and molecular research, a comprehensive explanatory mechanism for spatial memory has slowly developed. Part of the research consisted in separate investigations of different processes and parts, sometimes not even directly related to mice navigation skills. Another part of the explanatory efforts involved tying together these separate investigation results into an explanatory mechanism consisting of different levels. This is an important scientific task, as it helps to mutually elucidate the formerly separate results: temporal constraints of spatial learning can be explained with reference to LTP processes, medical intervention can be improved with knowledge of molecular processes, and so on (Craver 2002).

This again confirms the explanatory pluralism mentioned earlier in this part, implying that for a complex phenomenon like spatial memory we invoke – and integrate – theories pertaining to various components of the phenomenon that shed light on its various specific properties.[101] As a result, each theory has only a 'relative significance' (Beatty 1997). The integration of all relevant theories is a difficult part of research and it is here that mechanistic explanation can offer a useful methodology, maintaining as it does the necessary causal and theoretical pluralism.[102]

When providing a methodology for the integration of several theories regarding a phenomenon, it is useful to note that a mechanistic explanation is usually the result of a gradual development. In light of this, a distinction has been proposed between a mechanism sketch and its eventual schema (Machamer, Darden et al. 2000), or between 'how-possibly' and 'how-actually' explanatory mechanisms (Craver 2007). For example, presenting the explanatory mechanism of spatial memory as consisting of three levels of submechanisms, Craver admits that "this explanation into four levels is surely an oversimplification" (Craver 2007 169). Depending on the

[101] In other words, integration of different theories amounts to finding constraints at several mechanism levels that together determine the space of options for a phenomenon (Craver 2007).

[102] Wimsatt explains what often happens in a situation where pluralism is required: "given the difficulty of relating this plurality of partial theories and models to one another, they tend to be analyzed in isolation with unrealistic assumptions of system-closure (to outside 'disturbing' forces) and completeness (that the postulated set of variables includes all relevant ones)" (Wimsatt 2007 180). This is similar to the independence and uniqueness that are often and mistakenly ascribed to a classification, although pluralism is usually more appropriate in this context as well (Hacking 1991).

research question and available techniques, the coordinated activation patterns of hippocampal cells or interactions between different neural areas could have been added to the explanatory mechanism, or further inquiries into underlying molecular submechanisms could have been performed.[103]

Finally, for the integration of multiple theories into a comprehensive explanatory mechanism, investigations exploit its multilevel organization and interactions. Interlevel experiments are used both for a bottom-up investigation of a component's function at a higher level – what role does the NMDA receptor play in LTP and consequently in learning? – or for a top-down investigation of the submechanism responsible for a particular function – how does alcohol addictive behavior affect the mechanisms involved in memory formation and learning?[104] In either case, a particular intervention targets specific components at a particular level, while potential effects are detected at another level (Craver 2002). In addition to interlevel experiments 'looking up' for functions or 'looking down' for responsible mechanisms, researchers are also 'looking around' in order to ascertain additional components at the same level or relevant environmental conditions (Bechtel 2009b). In sum, not only does mechanistic explanation provide resources that can help in the integration of various scientific insights, it also assists in articulating modes for further investigating a phenomenon. Moreover, it does so without enforcing a particular metaphysical, reductionist, position.

Now that we have introduced and provided an example of mechanistic explanation, it is finally time to detail the heuristics guiding its development. It is in fact a gradual process facilitated by performing three different heuristics – not just once, but often iteratively. In the following three sections, we will shed light on the heuristics of definition, decomposition and localization. Together, they will clarify how in the next parts the insights from philosophy, psychology and neuroscience can be combined to provide an integrated account of the space of actions that a subject develops, learning to perform his actions in conformation with a wide variety of constraints including his own intentions. To prepare for the explanation of such a dynamic process, the sections below on heuristics will contain references to skill learning and a specific section will be devoted to dynamic modifications of an explanatory mechanism.

---

[103] Indeed, 'bottoming out' an explanatory mechanism at lower levels is relative (Machamer, Darden et al. 2000), and a decision about digging further may be guided even by considerations of its cost-effectiveness (Wimsatt 2007). In the next sections, it will become clear why the explanatory relevance of ever lower levels usually decreases.

[104] Alcohol affects i.a. the NMDA receptor, which we know is involved in LTP (Lovinger 1993).

5.3 Defining the phenomenon as a first step

In the context of previous explanatory approaches, we witnessed that definitions of psychological functions or computational tasks generally play a role. However, the role of definitions was quite different. In chapter I.2, we found that Bennett & Hacker argued for the consideration of a function's definition as the a priori delineation of its domain, without empirical, neuroscientific data being able to affect the definition (Bennett and Hacker 2003). Conversely, we found the suggestion stemming from the NCC approach to redefine consciousness in line with the neural correlate of recurrent processing (Lamme 2006). Not content with either approach, we repeatedly referred to the complex phenomenon of singing. By associating singing with similar yet different phenomena, like vocal signalling, infant crying and babbling, bird song, and expert song, we learned that it makes no sense to aim for a single definition of singing in light of these variabilities and dynamics. Since mechanistic explanation is considered to be well-equipped for the explanation of such variable and dynamic functions, the question to be answered is what role it leaves for a definition of the phenomenon under scrutiny.

As expected, defining does play a role in the mechanistic explanatory account as well.[105] As an explanatory mechanism is constituted by component parts and operations that together are responsible for a particular phenomenon, a faulty definition of the latter will have consequences for our ability to develop an adequate explanatory mechanism of it. Such a definition should give us a first delineation of the phenomenon that we are trying to explain. It should also help to limit our explanatory work, since a mechanistic explanation is specifically tied to a particular phenomenon: "boundaries of mechanisms—what is in the mechanism and what is not—are fixed by reference to the phenomenon that the mechanism explains" (Craver 2007 123). A definition helps as a delineation, but it can also misguide research.

Such misguided research can be blamed on several mistakes with respect to this initial step: "because one has tried to explain a fictitious phenomenon, because one has mischaracterized the phenomenon, and because one has characterized the phenomenon to be explained only partially" (Craver 2007 128). As our reflections on singing testify, this admonition to present correct definitions is not as simple as it seems. Plant song may be a fictitious phenomenon, but does that hold for baby song, too? Must we delineate singing from other social or communicative actions, or is that impossible?

---

[105] Strangely enough (Bechtel and Richardson 1993) fails to mention definition as a separate heuristic. This may have to do with the fact that most of their examples do not raise the kind of debate that we find regarding higher cognitive functions.

These reflections also point out the usefulness of the development of a taxonomy or classification as a first step in a research project. In doing so, a phenomenon is being related to neighboring phenomena and separated from others, providing both a first delineation and a first description of the relevant factual domain. As mentioned before, though, it is important to realize that in the life sciences it is generally not possible to provide a single classification for a domain of phenomena, but that pluralism reigns (Dupré 2001 ; Hacking 1991).[106] Even though it is possible to include each phenomenon into a plurality of classifications, such classificatory work can still help to guide further research by setting it into a variety of relations to other phenomena. In doing so, defining can be used as a heuristic, helping even the investigation of phenomena as odd as 'blindsight' (Keestra and Cowley 2011).

Nonetheless, defining and classifying can also go astray and then create specific problems: for example, if spatial memory is in fact an ill-defined phenomenon, it may be doubtful whether its investigations can consistently identify an explanatory mechanism (Sullivan 2010). Most common errors occur when distinct phenomena are erroneously lumped together instead of being assigned to different classes, or when two kindred phenomena are mistakenly split into two different classes (Craver 2007). Clearly, in both cases developing an explanatory mechanism will lead to serious issues: in the first case, it may be impossible to find overlapping components of the mechanism that can simultaneously explain two – actually completely distinct – phenomena. In the second, researchers may find that the explanatory mechanisms of the two phenomena overlap to such an extent that it is worth inquiring which distinctions have kept the two phenomena conceptually separate. In both cases, it may be necessary not only to scrutinize the explanatory mechanism(s), but also to reconsider the definition(s) at stake.

Consequently, defining and delineating a phenomenon and explaining it are intimately related to each other. Although it has been argued in (Bennett and Hacker 2003) that (re-)defining is logically separate from investigating a phenomenon, we hope by now to have convincingly argued that this cannot be maintained, as this position relies on – logically - untenable views of what definitions and explanations are (cf. also Keestra and Cowley 2011). Similar to this recurrent definitory work when developing explanations, it may happen that researchers revisit their initial

[106] See for instance the interesting analysis of genos and eidos as relative, not absolute, units in Balme's commentaries in (Aristotle 1972) and related issues in (Gotthelf 1987). The limited value of definitions in Aristotle's biological works is also discussed in (Gotthelf 1997). A modest attempt to relate these issues to mathematical definitions and Aristotle's awareness of the misleading role of language was presented in our (Keestra 1991).

decomposition of a phenomenon and subsequently 'reconstitute' it on the basis of gathered evidence (Bechtel and Richardson 1993).

In closing this section on the relevance of definition for the mechanistic explanatory approach, it may be useful to refer to an example other than bird song, as this dissertation will focus more specifically on human determination of action. We will find in the next chapters that different components constitute the complex mechanism that determines the action a human agent performs in a given situation. Interestingly, this mechanism is also very dynamic, being capable for example of various forms of learning and automatization. If we consider the listed properties of automatization, it already provides a clue for the comprehensive classificatory web that can be developed for defining such automatization. For automatized action has been defined as being largely: "unintentional, uncontrolled/uncontrollable, goal independent, autonomous, purely stimulus driven, unconscious, efficient, and fast" (Moors and De Houwer 2006 297). Obviously, this provides some clues for initial research – which is what a heuristic at least should do. Similar differences can be found between novice and expert song, some components having decreased relevance in the latter. For instance, an expert singer may no longer need to listen to others to keep his tone, or watch a metronome to keep time.[107]

After such a first definition or delineation of the phenomenon, two further heuristics are employed in its empirical research. While definition embeds a phenomenon in a wider web of relations, decomposition then helps to divide the complex phenomenon into smaller portions that facilitate research. Subsequent localization should then help to determine ever more precisely how a phenomenon or its components are performed by a particular organism or system.

5.4 Facilitating the explanatory task by decomposing the phenomenon

Notwithstanding corrigibility – or rather pluralism - research generally starts with a definition and delineation of a phenomenon, as we argued in the previous section.[108] Subsequently, its investigation is facilitated by dividing it into explanatory sub-tasks. This is done by applying the second heuristic, that of decomposition. The phenomenon is considered to consist of a number of components, which in their integrated

[107] In any case, melodic and temporal processing during singing are largely independent. Relations between specific components of song can differ a lot, including differences between production and perception conditions (Peretz and Zatorre 2005).

[108] In their influential book on 'discovering complexity', the authors present only the two heuristics of decomposition and localization (Bechtel and Richardson 1993). The preliminary delineation of an object appears not to be recognized as a first heuristic, even though it is a task that is different in kind from its subsequent decomposition.

unity amounts to singing, for instance. Remember that an explanatory mechanism consists of an organized structure of component parts and operations which together produce a phenomenon (Bechtel 2008). Decomposing the phenomenon means that several such components are being distinguished, which merit relatively independent investigations and thus facilitate explanatory research.[109]

However, similar to the pluralism pertaining to the classification and definition of a phenomenon, a pluralism of its decompositions is also possible. For instance, in the cognitive sciences phenomenal, mechanistic, functional, and anatomical decompositions are used in parallel for the development of an explanatory mechanism of a particular cognitive function (Bechtel 2008).[110] Obviously, each form of decomposition then focuses on different aspects of it. In case researchers are investigating a cognitive function in order to explain it, it seems plausible to offer at first a phenomenal decomposition.[111] Such a decomposition is usually based upon the observation of behavioral or verbal responses, like when different subjects perform specific tasks under different conditions or when patients are being studied.[112] In the case of memory, this has led to multiple phenomenal decompositions, for example to distinctions between short- and long-term memory, to declarative and procedural memory, and correspondingly to different forms of amnesia (Bechtel 2001a). Similarly and perhaps more familiar are the debates about different phenomenal decompositions of action, starting with Aristotle's notorious distinction between voluntary, non-voluntary and in-voluntary action in his Nicomachean Ethics.[113]

---

[109] Dennett presents decomposition of a computational task also as a way to develop artificial intelligence. Using a nice metaphor, he writes: "[e]ach homunculus in turn is analyzed into smaller homunculi, but, more important, into less clever homunculi.... being reduced to functionaries 'who can be replaced by a machine.'" (Dennett 1978 80). Lacking, however, in this description are the organization of and interactions between the homunculi, without which a decomposed task will not be complete. Note that interactions are also level-specific, with neurochemical interactions taking place at another temporal and spatial scale than the electro-physiological interactions even if the latter are partly composed of the former.

[110] For example, Aristotle was the first to establish a framework for a psychological science and proposing different subdomains or faculties. This has provoked criticism, since his proposed decomposition is held to be outdated (cf. Clark 1997 221; Vanderwolf 1998). The conclusion should be that a specific decomposition may at some point stand in need of correction or even outright rejection as a result of research, but not that we should do away with definition and decomposition as a heuristic entirely.

[111] Wimsatt discusses the descriptive complexity of an organism by aligning several types of its decomposition – e.g. in terms of its physico-chemical or anatomical compositions, its developmental gradients - and proposes a corresponding notion of complexity that refers to the different degrees of non-isomorphism of these decompositions (Wimsatt 2007, cf. figure 9.1, p. 183).

[112] In most cases a phenomenal decomposition, a starter for explanatory research, will be based upon a phenomenal description which is itself akin to a functional analysis: most cognitive phenomena do allow for such a functional analysis. Piccinini and Craver even state somewhat generalizingly: "Psychology should not content itself with the discovery of merely phenomenally adequate functional descriptions that fail to correspond to the structural components to be found in the brain. It should aim to discover mechanisms. To explain in cognitive psychology and neuroscience is to know the mechanisms, and explanation is what functional analysis has always been all about" (Piccini and Craver 2011 308).

A further decomposition, then, is the mechanical one that actually starts to focus on the components that together produce the phenomenon at stake. Although normally hard to distinguish or even perceive in a phenomenal analysis, most cognitive and behavioral phenomena are constituted by a host of component parts and operations. For example, evidence was gathered relatively early on that vision consists of components like color vision, motion vision, face recognition, while two decades ago already more than 30 different mechanism components were distinguished and explained (Felleman and Van Essen 1991). Regarding memory, and irrespective of its phenomenal component (short- or long-term, declarative or procedural, etc.), researchers have traditionally distinguished mechanism components like the encoding, storage and retrieval of memory (Baddeley 1976). Meanwhile, the nature and number of components have been modified, with researchers now seeking to explain additional components like consolidation, re-consolidation, activation and so on (Hardt, Einarsson et al. 2010).

As these examples underline, the results of applying the heuristic of decomposition are likely in need of modification as new insights are gathered. The need for modification will be diminished when the decomposition is the robust result of a variety of studies, and not just of a single type of study - the latter being often the case when a first decomposition of the possible explanatory mechanism is proposed. Exemplifying this is the influential decomposition of language processing, which at first depended heavily on the patient or lesion studies published by Broca and Wernicka. This decomposition suggested a rather simple distinction between speech production and perception (Bechtel 2001c), but has now been superseded by a much more differentiated and nuanced version, although speech production remains elusive partly because animal models are not available for its investigation (Hagoort and Levelt 2009), confirming the need for interdisciplinary efforts.

Even though we have to treat patient or lesion studies with great care, they are often helpful for the first attempt at decomposing a phenomenon's explanatory mechanism.[114] These are then complemented with animal studies, computational studies, experimental studies and so on. Importantly, researchers should not satisfy

[113] The difficulty of such decomposition is all the more evident when we realize that Aristotle provided only a twofold decomposition of action in his Eudemian Ethics, rejecting the differentiation of non-voluntary and in-voluntary action. Although Kenny considers the former as an inferior version (Kenny 1979), we beg to differ, as the former decomposition allows an observer to determine the voluntariness of an action even post factum when taking an agent's sorrow or remorse into account ((cf. Ethica Nicomacheia III, 1)

[114] Obviously, extreme care is required when generalizing from pathological studies to the explanation of normal subjects' cognitive and behavioral responses. It is far from clear, for example, whether double dissociations allow us to draw general conclusions, given the complexity and plasticity of the brain (cf. Karmiloff-Smith, Scerif et al. 2003 ; Orden, Pennington et al. 2001 ; Plaut 1995).

themselves with a decomposition based upon only pathological or, conversely, standard conditions. Instead, the decomposition of a phenomenon can be made more robust with the addition of: "descriptions of the multiple features of a phenomenon, of its precipitating, inhibiting, modulating, and nonstandard conditions, and of its by-products" (Craver 2007 128). However, particularly in the case of complex and dynamic phenomena, we may not always get the same decomposition results when investigating different phases of a particular function. This may be the case when we investigate 'nonstandard' cases of a particular phenomenon like in expert singing, or generally in automatized skills. The question is, how mechanistic explanation can deal with such dynamic aspects of a particular phenomenon. We will take up that question in section I.5.6 below.

Particularly with regard to the decomposition of the mechanism, research requires the iterated application of this heuristic, due to the mechanism's structure. As discussed in section I.5.1 and exemplified above with the explanation of spatial memory, an explanatory mechanism is considered to exhibit a hierarchical organization of interacting organized levels. This is not peculiar for mechanistic explanation alone, as Marr also considered such recursive decomposition.[115] In his view, neuroscientific research involved "the study of particular mechanisms, these being assemblies made from basic components" (Marr 1980 199).[116] A similar assumption guides mechanistic explanation, implying that each component of an explanatory mechanism in turn allows further decomposition in terms of its component parts, operations and their organization. Indeed, this assumption concerns the 'near[ly completely] decomposability' of complex systems coupled with their hierarchical organization (Simon 1962). Of course, when employing this heuristic iteratively, researchers must ask themselves whether further 'bottoming out' an explanatory mechanism is still relevant for their explanatory goals (Machamer, Darden et al. 2000) and whether

[115] The affinity between Marr and the mechanistic explanatory approach has generally been overlooked in the literature. Bechtel writes, for example, "[e]ntities at different levels of organization stand in a part-whole relation to one another, whereas Marr's levels of understanding involve different perspectives or modes of analysis directed at the same entity or process" (Bechtel 2008 25, n. 11). Craver likes to compare Marr's levels of analysis with the levels of realization in the sense of Kim, while distinguishing these from mechanistic levels (cf. Craver 2007 165). However, given that Kim considers his levels also along the mereological lines just like Craver does (Kim 2000), Marr may be disagreeing less with both Bechtel and Craver than they assume.

[116] Although this (fourth) level may not be a principal addition to his methodology – reason for its relative absence in Marr's writings – it does for us signal two relevant aspects. First, it emphasizes that Marr's methodology refers not just to the relation between different theoretical perspectives, but also to the nature of the object of cognitive science. Second, by proposing these two neural implementation levels, Marr also appears to assume that recursive decomposition is a necessary ingredient of research, suggesting to us a particular structure of the explanatory mechanism that may be its result.

continually expanding their research in this way is cost-effective (Wimsatt 2007).

In a converse direction and less common in cognitive neuroscience, a phenomenon can be taken as an non-isolated explanandum ready for decomposition. Instead, often a phenomenon also figures itself as a mere component in an overarching and complex phenomenon. In such a case, researchers are looking for the 'role function' of the original phenomenon in its wider context, which usually also helps to explain several of its properties (Craver 2001). This applies when researchers aim to account for the role function of spatial memory in mouse navigation, of singing or skill learning. However, not every single functional interaction between a particular phenomenon and other phenomena is a satisfactory basis for assuming that they are in fact components of a distinct and overarching explanatory mechanism. If spatial memory and navigation only co-occurred at rare and exceptional moments, for example, their being a part of a comprehensive mechanism would seem unwarranted.[117] However, if they always co-occur, neglecting spatial memory's role function likely impedes its explanation. These considerations will return in the next section when the third and last heuristic, localization of an explanatory mechanism, is at issue.

5.5 Localization of the decomposed phenomenon

After defining a phenomenon and subsequently decomposing it into components, researchers aim to find a 'locus of control' for the responsible mechanism or one of its components by tentatively localizing it somewhere in the system or organism that displays it (Bechtel and Richardson 1993). Localization in itself being common to scientific efforts, the mechanistic explanatory approach can help to further clarify its procedures.[118] Such localization efforts are meant to further determine the plausible options for a mechanism that is responsible for a phenomenon and to exclude regions from this 'space of possible mechanisms' (Craver 2007 247-248). Before we describe how such localization may be carried out, a reservation has to be made.

---

[117] This naturally raises the question whether this explanatory approach accepts the thesis of an extended mind (Clark and Chalmers 1998), of embedded or distributed (Hutchins 2010), or enactive (Di Paolo, Rohde et al. 2010) cognition. Generally, this approach is not unsympathetic towards it, allowing mental mechanisms to perform functional roles in overarching phenomena. However, as soon as the locus of control has to be identified, and spatial, temporal or processing constraints are to be specified, it will often turn out that an individual agent is relatively autonomous vis-à-vis his environment (Bechtel 2009a). It is doubtful whether the 'transparency constraint' suggested by Thompson is strict enough and complying with these other constraints (Thompson and Stapleton 2009). Indeed, claims for embedding or extending mental mechanisms can easily be overstated.

[118] Although one may wonder how it is possible that, after acknowledging the complexity of a function and its emergent properties researchers still believe they can localize a complex function in an organism, Wimsatt rightly notes that it is a: "howling non sequitur that functional organization is not physical" (Wimsatt 2007 190).

As one of the most thought-provoking phenomena in contemporary physics is the appearance of non-locality or non-localizability, assigning a prominent role to the heuristic of localization may cause wonder.[119] Similarly, many large-scale systems or dynamic interactions between systems also appear to withstand localization, if localization is taken to imply the reduction of systemic properties to the properties of their distinct constituting entities. Three responses to this reservation are in order. First, as declared above, mechanistic explanation is useful to integrate and organize explanatory insights that are of a pluralist nature. Yet it does not have the ambition to replace all other forms of explanation, as it is specifically fit for 'near decomposable' phenomena (Bechtel and Richardson 1993 ; Callebaut and Rasskin-Gutman 2005 ; Wimsatt 2007). Second, not only are many phenomena not 'near decomposable', even for phenomena that are decomposable it is most likely that at some level of decomposition we will meet components or operations that resist further decomposition and thus resist further mechanistic explanation. However, the lowest level of a particular explanatory mechanism is *not* even meant to be the same as a lowest level of reality (Bechtel 2008 ; Craver 2007).[120] Third, even if such non-decomposable components are found at the bottom of an explanatory mechanism, this does not imply that such explanation is reductionist in the common sense of the word. Given its emphasis on the organization and interaction between components, which add additional levels of new, emergent properties to the mechanism, the acknowledgement of non-decomposable and non-localizable entities does not contradict their involvement in phenomena that do allow mechanistic explanation (Wimsatt 2007).[121] Indeed, almost every phenomenon in the material world is a demonstration of this obvious yet often confused or misunderstood fact. Leaving this misunderstanding behind, let us now look closer at the heuristic of localization.

---

[119] Thought-provoking as the might be, the interpretation of the modal status of these appearances is far from settled. Indeed, Dieks argues that an empiricist, Humean interpretation has much in its favor (Dieks 2011).

[120] This is why reductionism within mechanistic explanation is qualitatively different from reductionism in the common sense. In Bechtel's words: "if we adopt the mechanistic account, in which the notion of levels is defined only locally, then we are not confronted with the prospect of a comprehensive lower level that is causally complete and closed" (Bechtel 2008 148). Moreover, as noted earlier, mechanism levels are relevant for the explanation of a particular phenomenon and have no general ontological status, in contrast to the levels implied by common reductionist views.

[121] In neuroscientific terms, the crucial role for organization and interaction implies that not only neural cells but also their connections matter when it comes to decomposition localization – grey and white matter both matter (Ross 2010). Techniques for the precise imaging of connectivity in the living brain have become available only recently, so insight is still limited. One proof of its relevance is evidence that disturbed – anatomical or functional - connectivity may be partly responsible for cognitive dysfunctions (Andrews-Hanna, Snyder et al. 2007 ; Courchesne and Pierce 2005 ; Minshew, Williams et al. 2009).

Localization of cognitive functions has been a common strategy since the prehistoric days of trepanation, when skulls were penetrated probably to alleviate pathologies like brain haemorrhage (Missios 2007 ; Verano and Finger 2009). In modern times, localization was employed by phrenologists like Lavater and Gall to directly associate neural areas with psychological faculties or capabilities, influencing the investigations of Broca and Wernicke. The general assumption that even complex cognitive functions are localized, in an undecomposed fashion at a single location in the brain, was already matter of debate in the 19th century (Barker 1995) and has by now been largely abandoned. Still, remnants of such simple localizations still pervade the neuroscientific literature and are being critiqued as the 'new phrenology' (Uttal 2001). However, when researchers aim to develop a mechanistic explanation for a particular cognitive phenomenon's components, a direct correspondence between such a comprehensive phenomenon and a particular brain area is not at issue (Bechtel 2002).[122] Moreover, as we will see, the use of localization as a heuristic can be valuable even if in many cases it will have only limited success.

Whether it is on the basis of only a phenomenal decomposition, or based upon a first decomposition of the responsible mechanism, localization involves the further investigation of a phenomenon's physical properties. In cognitive neuroscience, this implies the formulation of heuristic 'psycho-neural identities' (McCauley and Bechtel 2001). As the notion suggests, localization here implies both the heuristic identification of two different levels of analysis – here: psychological and neuroscientific levels – and simultaneously also the distinction between levels of the explanatory mechanism.[123] This is comparable with Marr's approach, who admitted a loose relation between his computational and neural implementation theories of a particular function. Moreover, Marr also recognized that the investigation of a function's implementation in a particular mechanism must be directed at several levels, as it consists of components that are assembled in an organized fashion (Marr 1980). This was evident from our example of the spatial memory of a navigating mouse, which was localized somewhere in its hippocampus, while subsequent research further localized relevant components present in that area.

Often, deciding between different options of a phenomenon's locus of control and

---

[122] This observation is also relevant for the discussion on double dissociation studies in neuroscience, cf. footnote 114 above.

[123] Compared with Marr's reservations against strict interdependencies between his computational and implementation theories (Marr 1982), mechanistic explanations explicitly aim to explore the potential constraints available between different kinds of levels in order to determine explanations. In so doing, the levels of processing, organization and analysis as distinguished in (Churchland and Sejnowski 1988) are also used in combination and not only separately.

about its further components involves an element of choice, without the availability of strict criteria for making that choice. Wimsatt has proposed to consider the robustness of a phenomenon or a component as a criterion. The more robust a phenomenon is, he argues, the more easily detectable it will be by independent investigatory methods, and hence the more explanatorily fruitful and predictively richer (Wimsatt 2007 63). Clearly, opinions can diverge on the degree of robustness or relevance of a particular phenomenon or component thereof, as we will see. Indeed, controversies in the field of cognitive neuroscience often depend on such divergences.[124] However, with growing evidence, researchers can usually localize a phenomenon or its components ever more precisely.

An initial step in such a localization effort is to 'segment a system from its environment', for example by investigating whether the systems displays the phenomenon in different environments or under variable environmental conditions (Bechtel and Richardson 1993). Spatial constraints are indeed among the most relevant constraints that help limit the space of possible explanatory mechanisms for a particular phenomenon (Craver 2007). This usually first involves a designation of its locus of control as a whole, with later refinement as information about the spatial constraints of its components and their organization are obtained. Similarly, temporal constraints concerning the order, rate, duration and frequency of relevant activities play a role in determining the plausibility of a phenomenon's locus of control, particularly when considered in parallel with the spatial constraints. For example, at what particular locations are activities observable, during which phase of a particular phenomenon (Craver 2007)?

If researchers agree on a particular (sub-)system as a locus of control for the performance of a phenomenon independent from its wider environment, subsequent research may then seek to localize it – or its components - ever more precisely in a particular part of it.[125] Just like the recursive decomposition of a phenomenon implies its having a hierarchical structure, most localization techniques in cognitive neuroscience assume that a phenomenon is produced by a hierarchically structured

[124] In his critique of mechanist explanation, Moss focuses particularly on Craver's work and its normative tenor (Moss 2012). Indeed does Craver not seem to realize the consequences from the non-rigid nature of the robustness criteria mentioned here. On the other hand, Craver does emphasize the limited nature of any explanatory mechanism, valid as it is only for a particular phenomenon (Craver 2007).

[125] In cognitive neuroscience, even such segmentation is often disputed. Indeed, currently it is much debated whether cognition should be considered to be not only 'embrained', but also 'embodied' or 'situated' in a broader sense (cf. discussions in (Anderson 2003 ; Barsalou 2010 ; Mareschal, Johnson et al. 2007 ; Niedenthal, Barsalou et al. 2005 ; Wilson 2002). Where embodiment will generally satisfy some spatial constraints as the body 'travels along' with a brain, this will be much more debatable for broader interpretations of situated cognition. The connection between a cognitive phenomenon and an environmental condition is much looser, obviously, which is reason for caution with regard to further assuming cognition's structural embeddedness in more comprehensive systems.

mechanism. Above in section I.5.2, we mentioned the three 'directions' of research appropriate for the multi-level structure of explanatory mechanisms, as investigations amount to looking down, up and around (Bechtel 2009b). We referred to 'inter-level experiments', that is: interference, stimulation and activation experiments. Such experiments intervene at a certain level of the mechanism and aim to detect the consequences of the intervention at another level – either top-down or bottom-up.[126] For example, spatial memory was localized in the mouse's brain, with components being determined at several levels: hippocampal cells, synaptic processes involved in LTP, NMDA receptors (Craver 2002, see the fig. in the previous section).

These inter-level interventions and detections must involve different methods according to the level at which they aim, because levels are occupied with different component parts and operations and because new properties emerge at each level due to the organized interactions of the components at lower levels (Bechtel 2008).[127] Given these level-specific properties, we can employ stimulus-response studies, imaging studies, single-cell recordings or pharmacological studies in order to detect memory responses, for example, or how singing behavior alters due to such local interventions. The results will pertain to localization at different levels, showing the involvement of brain networks, brain areas, particular cells, and specific molecules. Consequently, researchers aim to integrate not only spatial and local constraints but also other relevant constraints when determining a particular phenomenon's explanatory mechanism (Craver 2007).

Similar to its decomposition, the search for the locus of control of a cognitive phenomenon often also benefits from brain lesion studies in patients. For localization purposes, this technique is often not very reliable given the brain's plasticity, evident in reorganizations that occur in a disrupted yet dynamic neural mechanism (Buonomano and Merzenich 1998). Nonetheless, Broca's and Wernicke's

---

[126] As mentioned earlier, it is important to note that in a complex and dynamic mechanism, other than in a purely aggregative system, there will be many properties that emerge at higher levels. Such emergent properties may in turn have interactions with lower level properties of the mechanism and also allow forms of interaction with the environment that the lower level components and operations by themselves are incapable of (Wimsatt 2007, particularly part III).

[127] Investigative techniques are often specific depending upon the mechanistic level at which they are applied. Wimsatt explains this by referring to organizational levels in nature, which are occupied by "families of entities usually of comparable size and dynamic properties, which characteristically interact primarily with one another, and which, taken together, give an apparent rough closure over a range of phenomena and regularities" (Wimsatt 2007 204). Generally, therefore, reduction is not plausible, and even less so in the case of a specific explanatory mechanism, where levels differ as a result of the organization of components at each level, as Wimsatt argued already in his (Wimsatt 1976). Successful reductionism of theories that apply to a particular level is also much rarer than its notoriety suggests. With regard to interlevel theory reduction, McCauley concluded even that: "The history of science reveals no precedent for theory replacement or elimination in interlevel contexts" (McCauley 1986 197).

studies of lesion patients suggested a particular decomposition of human speech, but additionally allowed these neuroscientists to provide a first localization of specific speech components in particular brain areas.[128] However, as mentioned in the previous section, speech has meanwhile been recomposed or reconstituted on the basis of subsequent research, corresponding with different and more elaborate localizations of the explanatory mechanism for speech (Bechtel 2001c).

Indeed, as inevitable and important as localization efforts in cognitive science may be, localization hypotheses are likely to require revision continuously. Such revision usually corresponds with more detailed insights in the constraints of the phenomenon itself and its explanatory mechanism. An extra complication for localization efforts is the fact that cognitive phenomena are also malleable through development and learning. This is the case with spatial memory, as it is with skill learning, expert singing. The question is whether or not such learning modifies the mechanism that is responsible for such a phenomenon. And if the explanatory mechanism changes during learning, is its localization different for novices and experts? Clearly, we are once again facing the question whether novice and expert performance can be explained with reference to a single explanatory mechanism, or whether we should distinguish between mechanisms responsible for their performance. This issue will return in the final section, with some concluding reflections on mechanistic explanation. But first, we will consider how mechanistic explanation can accommodate such dynamic modifications.

5.6 Mechanistic explanation and mechanism modifying dynamics

In what follows below, we will try to sketch some forms of mechanistically explaining the alterations involved, for example, in learning a particular skill. Even though it is generally acknowledged that our mental mechanisms "are plastic mechanisms that develop and change as a result of experience" (Bechtel 2008 240) and that learning in the sense of Long Term Potentiation has been explained mechanistically (Craver 2007), a more systematic treatment of such development and learning in the common sense has not been provided in the literature on mechanistic explanation, as far as we know.[129] Such a treatment would involve the articulation of various forms of

[128] Some other examples of decomposition and localization efforts in cognitive neuroscience stem from the research of memory (Craver 2002 ; Craver and Darden 2001), vision (Bechtel 2001b), and action understanding (Keestra 2011 ; Looren de Jong and Schouten 2007). Bechtel considers vision research as an exemplar in the Kuhnian sense, or a model, for the development of mechanistic explanation in cognitive neuroscience (Bechtel 2001b).

modification of an explanatory mechanism under the influence of experience or learning processes. Below, we will list four different forms of modification and present some empirical illustrations.

But first, a more principal remark is in order. Obviously, not all modifications in a multi-level mechanism are equally relevant. For example, distinguishable dynamic changes at lower levels of a multi-level mechanism can occur with high frequency without affecting relevant changes at higher levels, like changes in behavior.[130] This has to do with the 'dynamic autonomy' of mechanistic levels, entailing that most "micro-level changes don't make a causal difference at the macro-level" of a system (Wimsatt 2007 218). Such dynamic autonomy is a challenge to reductionists who believe that all events at the lowest levels transpire into appearances at higher levels of a mechanism and that the latter have no relative independence.[131] Conversely, modifications at higher levels of a system may not always be easily detectable as well.[132] In this dissertation, for example, we will discuss the dynamic changes that take place in the mechanisms involved in human action determination. In some cases, development leads to a reduction of the complexity of this process that is not always detectable in human behavior. For now, let us continue the exploration of the dynamics that can modify an explanatory mechanism and subsequently discuss the role of component parts and operations, their organization, and the mechanism's interaction with its environment.

A first modification to be considered is related to the set of mechanism parts that

---

[129] Bechtel has in some recent publications elaborated on the inclusion of dynamic systems theory in mechanistic explanations of dynamic functions like circadian rhythms (Bechtel and Abrahamsen 2010), arguing that the two methodologies can complement each other (Kaplan and Bechtel 2011). Although reference will be made to this work here, it has not really touched upon the kind of processes that interest us, nor has it systematically investigated the possible modifications of an explanatory mechanism due to dynamic changes.

[130] As mentioned in footnote 48, Wimsatt makes this point regarding multiple realizability and notes that philosophers of psychology tend to overlook the prevalence of such multiple realizability (Wimsatt 2007). However, Bechtel warns that many observations of multiple realizability of cognitive functions fail in drawing an adequate comparison between the different instantiations of a phenomenon and particularly their functional characterizations (Bechtel and Mundale 1999).

[131] This dynamic autonomy of – particularly higher - mechanistic levels is partly due to the fact that organization forms in complex systems usually are relatively robust, making them less vulnerable to disruptions. At least two different forms of robustness organization can be found in evolved, developing systems: redundancy robustness – with several identical pathways in parallel - and distributed robustness – with a network of non-identical, alternative pathways (Felix and Wagner 2006).

[132] Simon notes in a similar argument that most systems are near-decomposable, making for an incomplete dynamic autonomy of their levels. Hence, our account of these systems will always: "fall short of exactness because the properties of the lower-level, higher-frequency subsystems will 'show through' faintly into the behavior of the higher-level, lower-frequency systems" (Simon 1973 25). As a result of this, the behavior of such systems are not so much law-like, but rather in the form of regularities that include some exceptions (Glennan 2008).

are involved in the explanandum phenomenon. Remember that we are not talking about the introduction of a new part into the organism or its brain – only the novel involvement of an available part into the explanatory mechanism for a phenomenon. Given the fact that neural 're-use' is common in the brain, we may expect alterations with respect to a mechanism's components as well (Anderson 2010). Particularly interesting for our purposes is the addition or deduction of components due to learning or experience, for example when LTP creates strong interactions between previously loose neural areas. Alternatively, a modular component can even emerge over time due to plasticity, when internal connections within a particular neural network are strengthened above a certain threshold.[133] Again, such a modification can develop while potentially leaving the phenomenon largely – particularly in standard conditions – intact.

The second modification will involve the operations that are performed by component parts of a mechanism. Again, LTP with its activity dependent alterations in synaptic responses and associated genetic, neurochemical and molecular processes (Bliss and Collingridge 1993) may figure here as an example of modified activation patterns. Just like learning in a subject often involves the modification of behavioral or verbal responses in specific situations, such learning is often constituted by dynamic changes of operations performed at lower levels of the explanatory mechanism. Depending on the rest of the mechanism involved, such a modification can influence a cascade of further mechanism activities, leading to rather novel behavior of the mechanism. However, the results may also be more modest, as when enhanced specific stimulus sensitivity of a component merely leads to increases of the speed and efficiency of a subject's responses.

Third, the organization of the components may be modified via development or learning. As mentioned earlier in these sections on mechanistic explanation, it is the organization or re-organization of components that is often responsible for the emergence of new properties within a mechanism. In a hierarchically – or rather: heterarchically[134] - organized mechanism, learning and experience can affect the configuration of the component parts and their interactions, thus altering a phenomenon. It can involve, for instance, the alteration or even the thinning out of the relevant organization, an intermediate mechanism component being skipped when

[133] Such 'modularization' as a result of learning and development will be treated more extensively in chapter I.2, which discusses neuroconstructivist accounts like those presented in (Karmiloff-Smith 1992 ; Mareschal, Johnson et al. 2007).

[134] See (Berntson and Cacioppo 2008 ; McCulloch 1945) and footnote 96 on the importance and prevalence of heterarchy.

direct connections between more distant components have developed. Alternatively, feed-back loops can develop, or two previously unconnected networks can become dynamically coupled, making the organization much more complex than in a serially organized or linear mechanism (Bechtel and Abrahamsen 2010). In this context, too, such complex organizational modifications will more likely obtain at lower levels of an explanatory mechanism and will not be directly mapped onto identical changes in the phenomenon to be explained.

A fourth and final mechanism modification that can occur is of a somewhat different character, as it involves more than just its internal components or organization. One of the results of increasing complexity of a mechanism and the emergence of new properties at higher levels is its expanding capability of interactions with the environment. Together with the increasing degrees of freedom that such a system owes to the emergence of properties at higher levels, there comes an increase in such interaction capabilities, as with the development of molecular compounds, with sensory systems, with locomotion, and so on. Put in simple words: "There should be more ways of interacting with a spouse than with a quark!" (Wimsatt 2007 223). Given plasticity and learning capabilities of neural mechanisms, these expanded interaction capabilities can also have a lasting impact on the internal composition of the mechanism.

It is important to note, however, that a mechanism's development and learning does not always lead to an increase of interactions. On the contrary, these processes can also yield new strategies for complexity reduction. This will be a topic in the next parts, where the process of 'sculpting the space of actions' will to some extent consist of such a reduction, as it involves an increased consistency and coherence of action – a welcome phenomenon when interacting with a spouse or in singing, for example. Similarly, learning often results in complexity reduction by reducing the number of dimensions of a content through foregrounding some of its dimensions to the detriment of other dimensions, for instance by chunking memorized contents – as will also be discussed in the next part (cf. Halford, Wilson et al. 1998 and commentaries).

These four modifications are the most prominent ones that can affect a mechanism performing a particular phenomenon. A few empirical examples may help to clarify these modification types. A prominent phenomenon discussed in this context is that of modularity, which is related to the modification of component parts and operations within an explanatory mechanism.[135] Generally, modules are considered to be – functional, if not anatomical - components which have some autonomy within a system or mechanism, performing a specific function, sometimes related

to a specific information domain (Barrett and Kurzban 2006 ; Mitchell 2006 ; Seok 2006). Interestingly, such modules often appear as the result of a process in which a mechanistic component develops through increasingly interacting neural networks, which as a result become increasingly specialized in specific operations and inputs (Karmiloff-Smith 1992). Such a modification also affects further operations within the mechanism, since increased interactions within a modular configuration of mechanistic components are usually associated with decreased interactions with external components – like with other brain components or with the environment (Meunier, Lambiotte et al. 2010).

Generally, therefore, we may say that in many cases a combination of different modification types will occur, as in the case of skill learning, where several modifications occur in parallel. fMRI studies of skill automaticity suggest that automatization relies to some extent on increasingly efficient neural interactions within an existing network, based upon a form of Hebbian learning. Such learning alone would not modify the explanatory mechanism in a far-reaching sense. Indeed, an additional and different type of change has been observed, involving the alteration of the neural areas that are being recruited during skill performance (Petersen, van Mier et al. 1998). This is due to the fact that automaticity of a particular cognitive or behavioral phenomenon in many cases involves a shift from deliberate action planning to direct stimulus-response associations (Graybiel 2008).

So it is not just an increasing efficiency of some operations at neural levels of the mechanism that is observable in experts, but also a recruitment of different neural areas, affecting higher levels of the explanatory mechanism. As usual in such complex mechanisms, other modifications are also observable – even if not directly relevant. Depending on the modifications, extra neural resources become available to experts. Consequently, experts can more easily perform an additional task without the skilled task being disturbed, while novices in such a case must exert extra control and recruit extra frontal areas (Poldrack, Sabb et al. 2005). The net result of these changes is therefore not just increasing speed and efficiency pertaining to the skill, but also an increasing capability of responding to other internal or environmental conditions, enhancing the flexibility of the expert. An expert singer, for example, is flexible in meeting the intonation problems of an accompanying instrument or a conductor's forgetfulness, while a novice singer is not.

As a result of these observations, we may well conclude with regard to complex

[135] In the next part, modularity will be discussed more generally. Let it be noted here that since (Fodor 1983), modularity has been much debated and has received many different interpretations, cf. the review in (Seok 2006).

and dynamic systems that apparent identity in cognition or behavior may well hide differences in the relevant explanatory mechanisms. Differences that are often undetectable when applying standard criteria but that can at other times play out and therefore raise questions. It is long overdue, for that matter, that the impact of cultural differences on brain and cognitive processes and their large-scale neglect in cognitive neuroscience is addressed (Arnett 2008 ; Henrich, Heine et al. 2010). An important issue would be whether enculturated brains differ only in their functional anatomy or perhaps even in their structural anatomy, making these differences more resistant to change (Han and Northoff 2008). For mechanisms in general tend to restabilize after having undergone modifications, especially as subsequent actions and developments build on these, as we will discuss in the next part (cf. Wimsatt 2007). It is therefore relevant to note that functional effects of culture-specific differences have been detected even in relatively 'simple' cognitive processes like perceptual information processing or direction of attention (Nisbett, Peng et al. 2001). Given the importance of perception and attention for environmental interactions, development and learning, the effects of such differences could be pervasive.[136] After all, whether it is for the learning of skills, memorizing of information, or even singing, perception and attention exert influence on these processes.

With these considerations of mechanism modifying dynamics, our long exposition of the mechanistic explanatory approach has nearly come to an end. Even though this approach combines several of the merits of the other approaches while adding some more advantages, in the next section we will observe that it has also some limitations, a number of which are neither new nor specific to it.

5.7 Some limitations of the mechanistic explanatory approach

After having presented the mechanistic explanatory approach rather extensively, let us not overlook some limitations or reservations that cling to this approach as well.

First, it was noted that this approach cannot escape from the difficult problem concerning the identity of a phenomenon and of its explanatory mechanism. Obviously, this should not surprise us, as this problem was already a concern in Aristotle's reflections on scientific explanation (Sorabji 1980).[137] As is evident, the problem will return at several phases during the development of a mechanistic explanation and not just surface with regard to the phenomenon's definition.

[136] Our earlier discussion of the distraction of attention from pain in section I.2.4 has taught us that even pain processing is sensitive to attention's functional role. As another example, depending on the attended feature of a set of stimuli, subjects will adapt their categorization of objects which will subsequently affect their interactions with these objects (Blair, Watson et al. 2009).

Indeed, it is the interdependence between all three heuristics and equally the interdependence between mechanism levels that can help researchers to cope with this problem. Forming a definition helps to decompose the phenomenon, mechanical decomposition may then receive confirmation or disconfirmation from preliminary localization efforts, which may have us revisit the definition, and so on. In the end, it is its overall robustness that supports a particular explanatory mechanism (Wimsatt 2007).

Second, it must always be kept in mind that an explanatory mechanism is related to a particular phenomenon and not a complete description of a system or organism that performs that phenomenon. This also implies that we have to be extremely cautious when generalizing the insights into a particular mechanism to another, neighboring, phenomenon. The constitution of a responsible mechanism for a cognitive function is not equal to the general constitution of the brain as particular components may operate in a different configuration when involved in another cognitive function, or: "levels of mechanisms are defined componentially within a hierarchically organized mechanism, not by objective kinds identifiable independently of their organization in a mechanism" (Craver 2007 191). This lack of generalizability of mechanistic insights is not equal for all levels, of course. The molecular interactions found in NMDA receptors are likely present in all such receptors in the brain, whereas the properties of hippocampal cells are different from other cells in that area: insights in components at higher mechanistic levels are mostly less generalizable, since those components have smaller prevalence and appear with greater diversity. The number of components in quantum theory is small, yet notwithstanding the huge numbers in which they appear, they always behave according to general principles. It is hard to maintain this for the relatively small number of individuals that make up global society.

A third reservation concerns the 'near decomposability' that is assumed by explanations that refer to multi-level systems. Simon did acknowledge that this assumption has both an ontological and an epistemological side to it: when systems are not decomposable, it will be hard if not impossible to explain their behaviors (Simon 1962). A similar position was taken by Marr, who also assumed a system to have a modular and hierarchical organization (Marr 1982). In the previous section, however, we saw that in some cases increasing decomposability is a matter of

[137] Indeed, since Aristotle it is a major problem how to present a being's integrated unity in its definition, which instead focuses on the being's distinctive properties, cf. (Kessler 1976). It is important to realize that Aristotle's interest in this issue is especially motivated by his interest in biological kinds and much less by an interest in Platonic, geometrical objects, even though many interpreters have overlooked this fact (Aristotle 1972).

development and learning. Indeed, when a neural network is not yet decomposable and localizable, its performance may be rather difficult to explain mechanistically (Bechtel and Richardson 1993). Generally, in cases where a system's modular and hierarchical organization breaks down and organizational homogeneity increases, when consequently levels are hard to distinguish and internal regularity decreases, we may need yet another methodology (cf. Wimsatt 2007 221 ff.).

This leads naturally to the fourth observation, that in any case at the lowest levels of each mechanism, recursive decomposition and localization come to an end. As was noticed above, our explanatory goals in cognitive science are mostly satisfied with insights in mechanistic levels above subatomic levels.[138] Delving deeper, researchers will inevitably reach components that are no longer decomposable. This may also hold for an omnipresent force like gravity, for which probably no explanatory mechanism can be presented. This limitation of mechanistic explanation with regard to components at a potential fundamental level does not hinder its applicability to most other phenomena.[139]

Finally, it must be emphasized that mechanistic explanation may have been developed as an alternative to deductive-nomological explanations, but law-like theories will still figure within a mechanistic explanation at many places.[140] More than anything else, it is the ability of the mechanistic approach to integrate interdisciplinary results of research that makes it suitable for the demands of cognitive science (Keestra 2011). Indeed, given our conviction that the explanation of human action must allow room for a causal and theoretical pluralism, mechanistic explanation's explicit acceptance of such pluralism is the reason for adopting it as a leading model of explanation in the following parts. Once again, this does not imply that all phenomena pertaining to human action allow mechanistic explanation.

---

[138] There have been several attempts at an explanation of consciousness – which for some has a surprising indeterministic aspect – with reference to quantum physical processes that take place in the brain (Hameroff and Penrose 1996 ; Koch and Hepp 2006 ; Libet 2004 ; Walter 2001). However, not only is it implausible that the specific properties of consciousness can be tied to the myriads of quantum phenomena in brain cells, it is also unhelpful to connect an ill-defined problem (consciousness) with theories that are hardly expressible in terms relevant to cognitive neuroscience (Segalowitz 2009 ; Smith 2009).

[139] Schaffer objects more principally to the assumption of a fundamental level and argues that the distinction of levels or hierarchical structures does not imply acceptance of this assumption (Schaffer 2003).

[140] Moss's critique of mechanistic explanation is partly directed against the contrast with nomological explanation that several proponents make. He underestimates the potential for combining the two forms of explanation, though. Furthermore, the following statement demonstrates that he overlooks the fact that an explanatory mechanism is relevant only for a particular phenomenon: "the presupposition of any functional, let alone mechanistic, analysis is the holistic assumption of a unified entity that acts flexibly and contingently to sustain its own existence" (Moss 2012 166).