

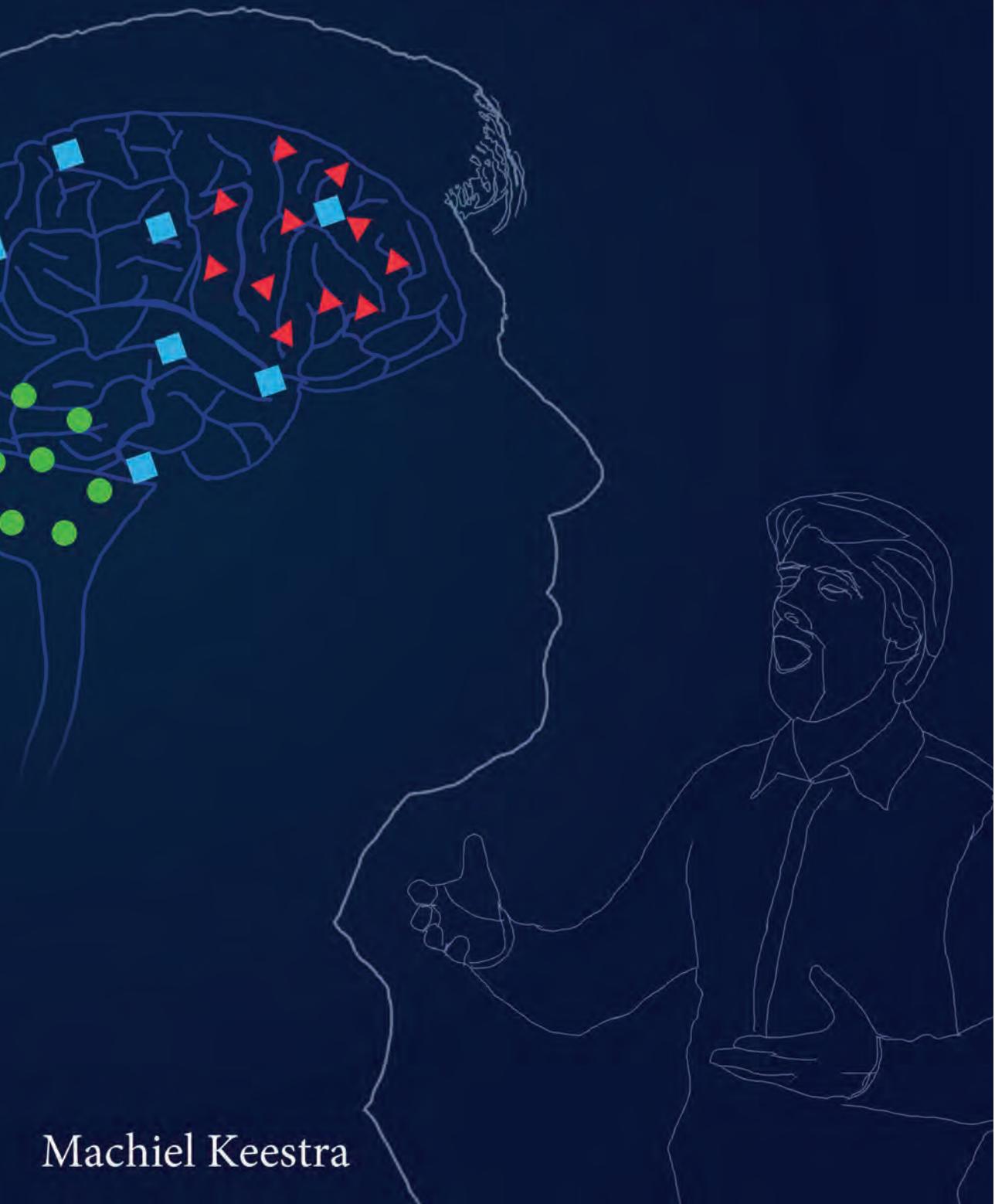


*Sculpting the Space of Actions. Explaining Human Action by Integrating Intentions and Mechanisms*

M. Keestra

# Sculpting the Space of Actions

*Explaining Human Action by Integrating  
Intentions and Mechanisms*



Machiel Keestra

ILLC Dissertation Series DS-2014-01.



## INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about and access to ILLC-publications, please contact  
Institute for Logic, Language and Computation

Universiteit van Amsterdam

Science Park 904

1098 XH Amsterdam

phone: +31-20-525 6051

fax: +31-20-525 5206

e-mail: [illc@uva.nl](mailto:illc@uva.nl)

homepage: <http://www.illc.uva.nl/>

This dissertation is open accessible at the Digital Academic Repository  
of the University of Amsterdam: [www.dare.uva.nl](http://www.dare.uva.nl).

# Sculpting the Space of Actions

## Explaining Human Action by Integrating Intentions and Mechanisms

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan

de Universiteit van Amsterdam,

op gezag van de Rector Magnificus

prof. dr. D.C. van den Boom

ten overstaan van een door het college voor promoties

ingestelde commissie,

in het openbaar te verdedigen in de Aula der Universiteit

op woensdag 15 januari 2014 te 13:00 uur

door

**MACHIEL KEESTRA**

geboren te Groningen

# Contents

<b>Introduction: intentional action and a sculpted space of actions</b>	<b>2</b>
How aristotle avoids the paradox of the expert by accepting causal pluralism	5
The importance of development and learning in aristotle's account	6
Conceptual innovation in aristotle's account of the development of expertise	9
Part I: investigating four methods of explanation in cognitive neuroscience	11
Part II: dynamic cognitive mechanisms and their stable adjustments due to development and learning	13
Part III: sculpting the space of actions for the performance of intentional expert action	15
<b>Part I – Conceiving and explaining: an intricate relation</b>	<b>21</b>
<b>1 Introduction: a common capability with divergent results</b>	<b>22</b>
<b>2 Concepts as delineations for empirical content</b>	<b>29</b>
2.1 Concepts as 'clean instruments' for neuroscience	30
2.2 Connective analysis and ascription criteria	33
2.3 Non-convergent and variable criteria, and their implications	37
2.4 Heuristic use of conceptual divergences, yet with limitations	42
<b>3 David Marr and the involvement of concepts in multi-level explanations</b>	<b>47</b>
3.1 The analysis of computations or tasks – not concepts - should guide scientific investigations	48
3.2 Constraints that co-determine the computations' appropriateness	52
3.3 Two further levels for multi-level explanations	55
3.3.1 <i>The algorithmic level and the representation of information</i>	55
3.3.2 <i>The implementation level and neuroscientific evidence</i>	58
3.4 A loose interdependency between levels	61
3.5 Modularity and some limitations of Marr's methodology	63
<b>4 Modest, all too modest: the search for neural correlates</b>	<b>68</b>
4.1 Identifying a minimal yet sufficient neural correlate – of what?	69
4.2 Further limiting the phenomenon and its correlates	74
4.3 Lessons on explanation drawn from the NCC research	77
<b>5 Mechanistic explanation and the integration of insights</b>	<b>81</b>
5.1 From the mechanistic world picture to the method of mechanistic explanation	85
5.2 Memory and the mechanistic explanation of learning	89
5.3 Defining the phenomenon as a first step	94
5.4 Facilitating the explanatory task by decomposing the phenomenon	96
5.5 Localization of the decomposed phenomenon	100
5.6 Mechanistic explanation and mechanism modifying dynamics	105
5.7 Some limitations of the mechanistic explanatory approach	110
<b>6 Concluding remarks after considering the four methodologies Dynamics of change and stability in cognitive mechanisms</b>	<b>113</b>

<b>Part II – Dynamics of change and stability in cognitive mechanisms</b>	<b>117</b>
<b>1 Introduction: from dynamics to stability and back again</b>	<b>118</b>
1.1 Kludges: mechanism adjustments and expansions	121
<b>2 Modularization as a process corresponding to learning and cognitive development</b>	<b>126</b>
2.1 Neuroconstructivism and the relevance of modularization	128
2.2 Modularization and the seven kludge characteristics	130
2.3 Modularization considered as a process of kludge formation	138
<b>3 Dual-process theories and a competition between forms of processing</b>	<b>141</b>
3.1 Distinguishing between forms of processing, irrespective of tasks?	144
3.1.1 <i>Considerations of the distinction between automatic and controlled processes</i>	148
3.1.2 <i>Processing limitations held responsible for the distinction between automatic and controlled processing</i>	150
3.1.3 <i>Memory systems invoked for the explanation of the distinction between automatic and controlled processing</i>	152
3.1.4 <i>Some strategies that allow a shift between automatic and controlled processing</i>	153
3.1.5 <i>Representational differences and the shift between automatic and controlled processing</i>	155
3.2 Automatization of controlled self-regulation and the seven kludge characteristics	161
<b>4 The brain as a mechanism capable of kludge formation and open to external information</b>	<b>166</b>
4.1 Symbols, simulators and the malleability and stability of cognitive processing	168
4.2 Simulators and the kludge characteristics	174
4.3 Reaching outside the skull: how can external objects become integrated?	179
4.4 Cognition-extensions and the kludge characteristics	188
<b>5 Dynamic mental mechanisms, kludge formation and establishing constraints on the space of options</b>	<b>198</b>
<b>Part III – Sculpting the Space of Actions with Intentions and Mechanisms</b>	<b>203</b>
<b>1 Introduction: multiple mechanisms yet stable patterns</b>	<b>204</b>
1.1 ‘Sculpting the space of actions’ – an important ingredient for the explanation of expert action	207
1.2 Determining an action via a cascade of intentions	216
1.3 The cascade of intentions and a sculpted space of actions	222
<b>2 Motor intentions: the first step in the hierarchy, or not?</b>	<b>226</b>
2.1 A philosophical analysis of motor intentions and guidance	228
2.2 Motor intentions and chunks: evidence about developing complexity at the bottom of the hierarchy	234

2.2.1 <i>Motor intentions, representations and the role of expertise</i>	237
2.2.2 <i>Motor intentions and mechanisms that change with growing expertise</i>	243
2.2.3 <i>Motor intentions and differential generative entrenchment of components</i>	247
2.2.4 <i>Motor intentions and consistency of action</i>	250
2.2.5 <i>Motor intentions and some evidence concerning their neural implementation</i>	253
<b>3 Proximal intentions: a mediating role</b>	257
3.1 A philosophical analysis of proximal intentions	259
3.1.1 <i>Resolving conflicts between action options</i>	260
3.1.2 <i>Proximal intentions and blocking habitual action</i>	263
3.1.3 <i>Proximal intentions and constraints for anchoring an action</i>	265
3.1.4 <i>Proximal intentions and their peculiar position in the agent's psychology</i>	267
3.2 Proximal intentions and cognitive mechanisms that determine anchored actions	270
3.2.1 <i>Processes for the resolution of conflicts between action options</i>	276
3.2.2 <i>Blocking a habitual action according to the CS/SAS model</i>	280
3.2.3 <i>Multiple processes and the constraints for anchoring an intention in a situation</i>	284
3.2.4 <i>Proximal intentions and some evidence concerning their neural implementation</i>	287
<b>4 Distal intentions: governing the intentional cascade?</b>	297
4.1 A philosophical analysis of distal intentions	303
4.1.1 <i>Closing the gap between distal and proximal intentions and sculpting the space of actions</i>	304
4.1.2 <i>Hierarchy and stability of a planning agent's web of intentions</i>	307
4.1.3 <i>A role for the imagination in weaving one's comprehensive web of intentions?</i>	309
4.1.4 <i>Narrative simulation as an additional resource of establishing one's agency</i>	312
4.1.5 <i>Narrative as simulation of distal intention, contributing to a sculpted space of actions</i>	318
4.2 Distal intentions and narratives and their reciprocal interactions with cognitive mechanisms	321
4.2.1 <i>From memorized experiences to the simulation of future actions</i>	324
4.2.2 <i>Distal intentions in the intentional cascade: from action control to mental time travel</i>	328
4.2.3 <i>Simulation and the flexible imagination of a future action</i>	332
4.2.4 <i>Narrative and additional benefits of the simulation of multiple distal intentions</i>	336
4.2.5 <i>The socio-cultural nature of some schemas for narrative simulation</i>	343
4.2.6 <i>Narrative simulation and some evidence for its implementation</i>	349
<b>Conclusion and Summary. Why sculpting the Space of actions matters</b>	356
<b>Figures: I, II, III</b>	371
<b>References</b>	377
<b>Samenvatting</b>	432

## CONCLUSION AND SUMMARY. WHY SCULPTING THE SPACE OF ACTIONS MATTERS\*

Why should we praise someone for performing so well, even though we usually reserve praise for consciously deliberated and chosen actions and less so for actions that appear to be produced automatically and effortless? Observing such action performance by an expert singer performing an opera role or a seasoned citizen engaging in moral action, one can easily fall prey to this paradox of expert action: instead of praising the expert, shouldn't we rather praise the novice, even though he may not be performing equally well and smoothly because he is at least deciding about his actions step by step? However, if we were to agree with this position, it would amount to admitting that during the process of acquiring expertise or skill, a person loses his – or her – admirability. For in proportion to his increasing expertise, his performance depends less and less on conscious and direct control of action. If indeed we take such immediate, conscious action control as a necessary ingredient of any form of intentional action, then we may be forced to withhold our expert singer the capability of intentionally executing his complex performance.

Such paradoxes have bothered philosophers, scientists and laymen alike since ancient times, trying to understand and explain human action. Indeed, Socrates aimed to avoid this paradox by positing that it is by definition through reasoning that an agent determines intentional and voluntary action. Aristotle clearly rejected this position, arguing against a simplified theory ('logos') that is at odds with our experiences and attitudes regarding this phenomenon and our reflection upon them (*Ethica Nicomacheia* 1145 b 27-28). In order to accommodate this, Aristotle added two elements to his account that allowed him to propose a more satisfying account. First, he recognized that human action is characterized by causal pluralism and not just determined by a single cause. Second, he realized that it may be necessary to carefully redefine current concepts or even to introduce additional concepts if paradoxes and inconsistencies arise within a theory of action.

It is such a navigation between conceptual and empirical insights that is undertaken in this dissertation, too. We wanted to do as much justice to the differences between expert and novice action as to their continuities. These differences are not only observable in the greater complexity, speed and flexibility of expert action in a given situation. In addition, an expert is generally better able to intentionally plan, organize,

---

\* On pages 371, 373, and 375, figures I, II, and III offer simplified representations related to the arguments made in Parts I, II, and III respectively.

modify and describe his action than a novice. Notwithstanding these differences, an expert didn't change his brain or body so we must explain how development and learning have enabled the same body and brain to produce a strikingly different performance.

For this explanation we have introduced and elaborated on the concept of 'sculpting the space of actions' as an explanatory tool (see section III.1.1). This concept allowed us to develop a comprehensive integration of interdisciplinary insights in (the emergence of) expert action, particularly philosophical and cognitive neuroscientific ones. In order to understand and explain how an agent determines his action in a given situation, we propose to consider it as a problem of finding a suitable candidate from the large number of actions that he could potentially perform. We propose to represent all of his action options as separate locations or subdomains within a multidimensional 'space of actions', specific to the agent. Each action option is located somewhere in this space of actions, its specific location being defined by numerous factors. Some action options are represented more prominently than others, occupying a larger sub-space at a more central location in the space of actions and therefore having a bigger chance of being selected and performed.

The space of actions that each agent has is not static. Instead, we argued that it is 'sculpted' in several ways, with both long-term and short-term effects. Long-term and stable changes that happen to an agent's space of actions are results of his development and learning. Such changes obtain when new actions are added to it, when well-practiced actions gain in prominence within the space, when unlearned actions are relegated to small and peripheral locations. Due to these long-term changes the options are no longer uniformly distributed in the space of actions. Instead, the experienced agent's space of actions is constrained in many different ways instead. According to this concept, when an agent is acquiring a skill or is gaining in expertise, his space of actions is subject to a sculpting process that affects particularly the sub-space of actions belonging to the domain of expertise.

We argued that when explaining an agent's action performance in a given situation, we should acknowledge that this process of sculpting the space of actions occurs in the short-term as well. For even though an expert's space of actions is sculpted more than a novice's, the selection of a particular option for performance is still subject to the conditions of the particular situation he finds himself in. A mountaineer who fell in sea will be less inclined to climbing than to swimming, an opera singer must do his best to act Don Giovanni-like charming to a detested colleague: external and internal conditions further sculpt the dimensions and structure of his sculpted space of actions in a more transitory sense during the action itself.

This concept of ‘sculpting the space of actions’ as an explanatory tool resulted from the preceding investigations made in this dissertation. First, in Part I, we critically discussed different methods of explanation used in cognitive neuroscience, looking for an explanatory method that can integrate both the causal pluralism and the effects of development and learning in an account of expert action. In Part II we then proceeded by applying the selected explanatory method to different theories of development and learning, accounting for both stable and dynamic effects of gaining expertise more generally. Finally, in Part III we turned to the explanation of human action and expert action in particular. Building upon our insights regarding explanation and about development and learning, we integrated philosophical and cognitive neuroscientific insights in it. This integration was facilitated by adding the concept of ‘sculpting the space of actions’ as a valuable explanatory tool. In the remainder of this section Conclusions and Summary, we will concisely travel after these navigations.

Part I was devoted to a discussion of four different methods of explanation pertaining to the field of cognitive neuroscience. All four methods offer solutions to how we should gather and integrate insights from neurobiological, computational, cognitive psychological and related studies in such a way that they together allow understanding and explaining complex phenomenon like intentional action. Complicating factor is that cognitive phenomena have proven difficult to define and without a clear definition it is unclear whether presented insights do in fact apply to the same phenomenon. Philosophical analyses can help with such conceptual matters but the four explanatory methods showed that they propose quite different relations between conceptual and empirical investigations.

The method discussed in chapter I.2 prescribes a crucial role to the way philosophers carry out conceptual analysis of a psychological function like consciousness or emotion. Its authors, Bennett and Hacker, maintain that empirical studies can only be usefully done on the basis of a clear definition that is reached through such an analysis. We pointed out how they assume that even for a notably complex function like consciousness, it is possible to develop a consistent conceptual framework, allegedly based upon an analysis of the concepts commonly applied to it and the behavioral criteria associated with it. This assumption was found to be unwarranted with regard to consciousness, or regarding a conceptual divergence like ‘blind-sight’ into account. Instead of rejecting such surprising concepts, as the authors propose, we defended that they be used as heuristics pointing us the way to unexpected interactions between functions or to a causal pluralism that has gone unnoticed. In other words, we argued

against strictly separating conceptual and empirical studies and for using insights from one as a constraint or heuristics for the other's investigations.

More productive is the method proposed three decades ago by David Marr and influential in cognitive neuroscience ever since, treated in chapter I.3. We found that it concurs to some extent with the previous one in that it assigns an important role to what is called the 'computational theory' or task analysis pertaining to a function, like vision. This computational theory should provide us with insights in the function's goal, taking into account also the function's role for other functions or in a wider context. This method strongly diverges from Bennett and Hacker's in that it explicitly prescribes how scientists should develop two more theories to gain a more comprehensive insight into a function. The 'algorithmic theory' explains how the information used for a task is represented and transformed, with usually several options available. Although Marr maintained otherwise, we found him using all three theories to constrain each other. For example, a particular task can theoretically be carried out with different kinds of representations, yet based upon the brain's neural properties one kind is more probably used than another kind of representation. It is by such an integration of insights, applied to its various objects, that cognitive neuroscience can make progress, as we defended throughout this dissertation.

Chapter I.4 was devoted to the method of explaining consciousness by looking for its neural correlates. We pointed out that this method does neither require a conceptual analysis nor a task analysis as it accepts that there is no generally accepted definition of consciousness available. A similar permissiveness was found with regard to its expectation that a particular conscious state should be 'mapped onto' a particular neural state, without prescribing the sort of relation between the two. However, notwithstanding their liberal stance, we found that researchers still cannot avoid differentiating between studies by using – sometimes implicit - concepts of consciousness. Alternatively we found how they created coherence between studies by looking for overlapping neural correlates, assuming that these findings do indeed pertain to the same object of study. Moreover, a particular neural process was presented as a defining criterion of consciousness. However, it still remains to be determined how this neural process contributes critically to consciousness, which is impossible without at least a preliminary definition or task analysis of consciousness. Determining the contribution of a neural process to it would then require formulating a computational theory or an algorithmic theory in Marr's terms for it, so we argued. Thus, investigating the neural correlates of a particular function in a fairly liberal way may indeed be useful, but only as a first step.

Chapter I.5 finally argued that the method of 'mechanistic explanation' facilitates

the required integration of insights better than the methods discussed so far by dividing the task of explanation of a function over many different perspectives and offering the means for their integration. It requires the application of a few heuristics for this task division and enables scientists to reconsider and adjust the formulated 'explanatory mechanism' in light of subsequent results. These heuristics are: the definition of a cognitive function, its decomposition in component functions, and finally the localization of these component functions in the organism and its brain. Each of these steps can be iterated in light of newly gathered insights or applied to further subcomponents. Memory, for example, has been defined as not just the storage but also the retrieval of information, as studies show that these can be differentially influenced or lesioned. With such a redefinition, the decomposition of memory has also changed and consequently, additional localizations in the brain have been scrutinized. We noted that developing a mechanistic explanation for a (component) function also benefits from formulating the three theories prescribed by Marr: what is the task of this particular function, what representations and transformations are involved for it, and how is it neurally implemented?

In addition to the fact that mechanistic explanation enables the integration of different insights, it is the only method that provides the resources needed for explaining the effects of development and learning which is particularly relevant for our project. For this, we developed here four different kinds of modifications that an explanatory mechanism can undergo, affecting the number and configuration of its components and also the interactions with its environment. Even though we acknowledged some limitations of this method of mechanistic explanation, we concluded that this method was most useful for the task at hand: explaining human expert action as being produced by a complex interaction between mechanisms and intentions.

Part II shifted to discussing several cognitive (neuro-)scientific theories about development and learning. Its aim was to consider whether we could apply the method of mechanistic explanation to these theories. It started with a preliminary general observation that development and learning generally lead to structural and stable changes in a mechanism responsible for a particular function. Because of their stability, such changes can accrue as they build upon previous changes, contributing to the hierarchical structure that complex and dynamic mechanisms usually have. As a result, earlier changes tend to become ever more deeply 'generatively entrenched', in Wimsatt's words, in the mechanism that subserves a particular function: a change has stable effects on the responsible mechanism and these effects are subsequently

involved in its further developments. We referred to such changes as cases of kludge formation, affecting both the structure and workings of the mechanism. Seven general kludge characteristics were set out, some of which appeared to be useful in our subsequent discussions of the theories of development and learning. Important, for example, was that as mechanistic explanation aims to elucidate an observable function, kludge formation must initially be characterized in functional terms. From this functional characterization we can unfortunately not directly derive a specific algorithmic theory nor a specific theory about its neural implementations, as was noted earlier. Indeed, it may be possible that differences between individuals can be found with regard to the representations or neural processes involved, even though these differences do not show up in their performances. A final kludge characteristic referred to the integration of environmental information in a function's explanatory mechanism. This explains why cultural differences can have a stable impact on it and not just on observable performances.

The first theory of development and learning, discussed in chapter II.2, was neuroconstructivism. Although focusing primarily on Karmiloff-Smith's work, which distinguishes between the stages involved in the acquisition of skills and expertise in children, we also applied this theory to adult learning. We found that neuroconstructivism assigns an important role to the process of 'Representational Redescription' that is involved in learning, concurring with the importance of algorithmic theories in cognitive neuroscience. During learning, the representations involved in executing a task do not remain the same but gain in complexity and structure, becoming increasingly available to the learner for explicit adjustment and correction, as when a singer learns to fathom the structure of his music score. Next to this process of 'explicitation', learning is also observable in the 'proceduralization' that accompanies it, affecting the task as it gets automatized and allows for less conscious control. This holds for our singer when he can sing a difficult score by heart. In that case, an expert can expand his performance by adding further elements to it or further refining it. As the term 'neuroconstructivism' suggests, this theory entails that during learning, the underlying neural mechanism changes by developing a more complex, modular structure. We argued that this 'modularization' concurs largely with the 'kludge formation' that according to us tends to affect mechanisms. We emphasized another insight from neuroconstructivism, which is that as a result of learning, there are several representations available to an expert for the performance of a task and not just a single one. Important for the present context is the consequence that an expert can be distinguished by his capability of switching to different modes of processing when performing a particular task, which a novice cannot do.

Differences between processing modes are what inspired a set of ‘dual-process theories’, the topic of chapter II.3. These theories distinguish between an automatic and a controlled mode of processing, differing among other things with regard to the information load they can process, the involvement of conscious control and the role of explicit knowledge. It implies that an agent gradually acquires the capability of performing a particular task in both modes of processing, as automatic processing is a result of his experience. This can be problematic for an agent because automatic processing can yield results that are stereotypical, for example, and not always in line with his performance in the other, controlled mode of processing. A singer performing Don Giovanni rather automatically may have difficulty avoiding a macho comportment, for example. We argued that such automatic processing is in itself beneficial for expert action, the important question being whether an agent can somehow control when his performance relies upon automatic processing or when it does not. Our discussion confirmed that some control is indeed available to the agent, pertaining to various aspects of his task performance. Regaining some control can be done by reducing the complexity of the information that is processed during the task, by changing its representation or by chunking it. Some self-regulation is possible, too, as when the agent somehow prepares for the conditions under which automatic instead of controlled processing would prevail. We argued that even such forms of self-regulation can lead to kludge formation and become integrated in the mechanisms responsible for automatic and controlled processing. So while admitting that task performance can rely upon different modes of processing, we rejected a strict separation of the two. Sculpting the space of his actions also implies that an agent improves upon his capability of regulating the different processing modes and the mechanisms involved that are responsible for his performances.

Chapter II.4 focused more specifically on a discussion of how external information becomes integrated in a mechanism responsible for an agent’s expert action due to learning and development. Especially as humans often rely on representations that employ language or symbols when they are learning, practicing or adjusting a task performance, the question is whether kludge formation obtains. We argued that this is indeed the case. Adopting Barsalou’s simulation theory, we explained expertise in terms of learners developing many ‘simulators’ that facilitate expert performance in a particular domain, like the domain of opera performance. A simulator consists of a complex, hierarchically structured network of component representations for a domain, which are stored in a distributed fashion across the brain and can be employed by different functions alike. Explicit representations and linguistic concepts can influence the formation, configuration and activation of these

representations. So when an expert action is performed, the agent in a sense ‘re-enacts’ a previous experience or action, or he composes a novel one by employing his stored representations. An expert therefore has multiple advantages compared to a novice, as he can employ a sculpted space of actions and has expertise in its targeted use. Hence, learning a new opera role is easier for an expert than for a novice.

We continued this chapter by discussing the theory of extended cognition presented by Clark and Chalmers. This theory holds that some cognitive or behavioral tasks rely so much upon the properties of external tools or other objects, that we should even include these in the mechanistic explanation of such a task. We argued instead not to expand the responsible explanatory mechanism by including external objects in it, but to explain the amazing interactions with objects by way of the human capability of developing complex representations in which object properties are integrated. Such a representation can then affect the mechanism responsible for a task. In other words, we aligned the simulation theory and the theory of extended cognition by applying our methodological insights.

In sum, we showed in Part II that explaining how an agent can learn to perform an expert action like performing the role of Don Giovanni should indeed be done by using the explanatory ingredients prepared in Part I. Development and learning, so we concluded, can be understood in terms of changes that affect relevant mechanisms and representations. The result is a complex situation, as an expert can perform a certain task in more than just a single way, for example via automatic or controlled processing or by employing one or another task representation, which a novice cannot. It is thanks to the process of sculpting the space of actions that an expert finds himself in that comfortable position.

Part III is devoted to a more specific investigation of intentional action, applying the methodological resources prepared in Part I, and the insights regarding development and learning from Part II. Indeed, we demonstrated that the explanation of intentional action is comparable to the explanation of expert action. Surprising as this may seem, by navigating between conceptual analyses of the components of intentional action and their empirical study, we demonstrated that an agent can only consistently perform actions according to his intentions when he has been sculpting the space of his actions. Part III started with a chapter expounding the framework to be used when discussing action intentions. Next, consecutive chapters are devoted to these intentions, always navigating from philosophical analyses to a discussion of empirical studies.

In chapter III.1 we introduced the notion of ‘sculpting the space of actions’, which

was mentioned above. We clarified in section III.1.1 why it is valuable to explain a given task as a problem in finding an adequate option in a so-called search space. It particularly facilitates such an explanatory effort as it enables the integration of multiple determining factors by representing each factor as an extra dimension to this multi-dimensional space. Expertise, we argued, should accordingly be conceived as a sculpting process affecting in several ways this space as Frith has done in the context of a language-processing task. Extending his analysis, we distinguished both a long-term and short-term sculpting process, having stable and dynamic effects on several related tasks. For example, a novice with a less sculpted space will be far slower and less adequate in his responses, but also in his perception and understanding of novel situations because a sculpted space is being employed by several cognitive processes alike.

These insights concerning a sculpted space will be integrated with Pacherie's framework containing three different levels of intention: motor intentions are responsible for guiding ongoing motor movements, proximal intentions for anchoring an intention in a given situation and distal intentions are the long term decisions about future actions. In section III.1.2 we described this framework and showed how it understands and explains intentions by integrating philosophical analyses from Frankfurt, Bratman and others with cognitive neuroscientific insights from Jeannerod and others. The framework organizes the different levels of intentions in a hierarchical structure and together with their interactions these enable an agent to eventually realize in motor movements a complex action that he decided to do long before the appropriate situation occurred. In this 'intentional cascade' framework, action representations were again found to play a central role, inviting their integration in a multidimensional, sculpted space of actions. Having laid out these notions of the intentional cascade and the sculpted space of actions, we then started with the discussion of the lower level of the cascade: first in a section with a philosophical analysis, second in a section regarding empirical cognitive neuroscientific insights.

Section III.2.1 contained a philosophical analysis of why motor intentions are distinguished which guide ongoing body movements. The fact that actions are continuously, fast and minutely adjusted to internal and external conditions suggests that these motor intentions play a role by integrating information about action goals, movements and a changing environment. Frankfurt was found to underline that we can observe how an agent continually receives feedback about his action and adjusts it accordingly. Such adjustments occur because an action, being different from a mere reflex, must be taken to stand in a particular relation to the agent's overall identity as cognitive, affective and attitudinal processes have determined it, all contributing to

consistency in his actions – even at this level of motor intentions.

Section III.2.2 then discussed empirical studies of motor intentions from Jeannerod and others yielding results that suggest how in fact such determination and guiding of an ongoing action is implemented. A motor intention guiding an action is constituted by a motor representation in a non-conceptual format that integrates, promptly and without conscious control, not just information concerning muscular movements but also information concerning the environment and the affordances for action that it contains. Experience was found to influence these representations in several ways, sculpting the space of an agent's actions.

Section III.2.2 continued with reference to De Groot's seminal studies of experiments with chess masters, elucidating how expertise affected the representations involved in their expert actions. Sculpting the space of their actions, they were found to assemble a very large number of increasingly complex and – hierarchically – structured representations, facilitating simultaneously their expert perception, decision-making and actions in complex situations. Interested in the representational redescription involved, we discussed the 'template theory' developed by Gobet and Simon which explains why experts are not only capable of handling complex situations but also flexible in doing so: their representations consist not just of complete chunks of information but also of complex templates with free slots that remain open for variable information. We pointed out how corresponding to these changes in representation, two neural processes obtain during learning of expert action, affecting subserving mechanisms. At first, expertise implies an increasing efficiency of neural activations during task performance, second, co-activations obtain which are due to other neural representations or processes related to the task at hand. Besides, it was found that specific neural areas or even single neurons can represent specific components of these motor representations, which are employed not just for motor actions but also for other tasks. We concluded that learning does indeed lead to the generative entrenchment both of particular components of the mechanisms responsible for the guidance of an expert's actions and of the specific motor representations involved.

In section III.3.1 we offered a philosophical analysis of proximal intentions. These fulfill fulfill a mediating role between the motor intentions and the distal intentions by specifying an action intention in a motor intention format, even though the action intention is initially made in a conceptual format long before an appropriate situation presents itself. So an expert singer somehow has to anchor his practiced interpretation of Don Giovanni's arias and behavior in a situation with specific stage props, ongoing directions of the conductor, other singers and so on. We applied to this the explanatory tool we introduced earlier, considering such anchoring as the singer's

further sculpting his stably sculpted space of actions. Proximal intentions' mediating role is particularly evident when an agent blocks a habitual action in an exceptional situation or when it is overridden by another – conflicting – distal intention, according to Bratman. In such cases, constraints are derived from the more comprehensive web of intentions and action plans that an agent typically has, about which more later. Together, these contribute to the consistency of his actions that is even visible in expert motor actions, in defiance of the paradox of action.

In III.3.2 we investigated empirical studies regarding the implementation of proximal intentions and argued that its intermediating role likely involves not just one but two distinguishable processes. We scrutinized the model of Norman & Shallice and colleagues which enabled us to explain both the habitual nature of complex actions by an expert and the potential modification or blocking of a habitual action. According to this model, large knowledge structures or action representations play a central role. With the interaction between a 'contention scheduling' process involved in composing the representation for a habitual action, and a 'supervisory attentional system' that can modulate or intervene in that process, we succeeded in explaining various properties of a proximal intention. According to this model, action representations are composed of many loose components that are put together in a hierarchical organization. This assembly of an action representation depends on the interactive activations with which components are related to each other and with other features like environmental triggers or goal conditions, as a result of development and learning.

We continued section III.3.2 by explaining how a sculpted space of actions is characterized among others things by specific interactive activation patterns. Effects of these can be observed when the opera singer manages to switch quite effortlessly during a rehearsal from Don Giovanni's seduction scene to performing Saint François's dialogue with the birds. Upon the activation of a particular action component – for instance by hearing the introduction of an aria – associated component representations are activated pertaining to other aspects of the performance. Other actions are facilitated, different anticipations regarding the environment arise and other constraints that depend upon his distal intentions are activated, too. We explained the importance of preparation and practice of such complex actions helps, as it increases the interactive activations between action components and can modulate these in a targeted way. Applying our notion of how development can involve the formation of kludge within a responsible mechanism, we explained how an expert is able to anchor and specify his intentions so fast, flexibly and consistently in contrast to a novice. We finally touched upon the neural

implementation of proximal intentions before turning to distal intentions, posited at the highest level of the intentional cascade.

Motor intentions and proximal intentions were shown to play indispensable roles in the performance of intentional action, doing so with relative autonomy. Yet they were also found to be only indirectly or partly related to the distal intentions, even though the latter are usually considered to be genuine intentions. Moreover, within the intentional cascade, not just top-down but also bottom-up influences are at stake, suggesting that distal intentions are themselves also influenced by the contents of an agent's prominent motor intentions, notwithstanding their different representational formats. Addressing these and other issues, chapter III.4 offered an extensive account of the roles of distal intentions, their implementation in the form of imagination or narrative simulation, the socio-cultural nature of schemas involved in such simulations and hypotheses about their neural implementation.

Section III.4.1 started by discussing Bratman's philosophical account of distal intentions, arguing that they play an important role in the complex task of coordinating and organizing an agent's actions. Without distal intentions constraining his space of actions, an agent likely engages in counterproductive actions, is incapable of realizing complex and temporally extended actions, and must keep cognitive resources free for continuously forming his intentions. We argued that expert action would be impossible under those conditions. This also implies that these distal intentions should not be reconsidered or changed lightly, but provide stable structure to an agent's sculpted space of actions. We concluded, however, that this account unfortunately has little to say about how an agent can represent the complex web of all of his intentions, which would enable him to organize and coordinate his actions and intentions comprehensively.

For that reason, we continued the philosophical analysis of section III.4.1 by proposing to remedy shortcomings of this account of distal intentions by enlarging it with Ricoeur's theory of narrative configuration of action. This theory contends that agents always engage in narrative, making action configurations that extend beyond the contents of single distal intentions to three further hierarchical levels: first, the level of socially shared practices, second, the level of plans regarding family life, professional life, and the like, and third, the comprehensive level of the unity of a life. By configuring and reconfiguring his narrative, an agent integrates heterogeneous ingredients like action components, goals, values, and temporal structure, but also environmental conditions and chance. Important to note is that this complex task is influenced by configurations or representations that a culture or tradition provides, even if an agent inevitably deviates from such examples. By way of his narrative, an

agent can not only describe his – past, present and future - actions and intentions, but also explain and perhaps adjust them and thus develop his identity as an agent, for himself and others alike. We argued that through narrative, an agent at least has the resources to plan and coordinate but also to evaluate and weigh his distal intentions in a way that Bratman nor Pacherie's intentional cascade were found to present.

In section III.4.2 we proposed to consider the implementation of distal intentions in cognitive and neural processes along the lines of a simulation theory, similar to the one discussed earlier in chapter II.4: action representations are stored not as a whole but as a hierarchically organized network of component representations throughout the brain, which are employed by different cognitive functions. Repeated employment strengthens the connections between components of a representation, causing the representation to become more deeply entrenched and more likely to influence future tasks. The simulation theory presented by Schacter and others can explain distal intentions and narrative in terms of 'constructive memory' processes. In doing so, the theory integrates both representations and mechanisms and confirms the notion of a sculpted space of actions.

We continued section III.4.2. by discussing options for the neural implementation of narrative, recognizing its comprehensive task in coordinating, organizing and evaluating the agent's actions and intentions. We argued that the default-mode-network discovered by Raichle and others is a candidate, as it was found to be involved in maintaining and evaluating information and to be activated during interpretive and predictive tasks, including those that are self-referential in nature. Doing all this, it plays an additional and important role in sculpting the space of actions of an agent, enabling him to become an expert who consistently and coherently performs actions that comply with his intentions to the surprise of observers.

What have we reached with these navigations and why did they have to be so extended? At the outset of this section, we pointed out how observations, conceptualizations, investigations and explanations cohere intimately with each other, making the task of explaining human action complex. We intended this study to contribute to the necessary integration of intentions and mechanisms for the explanation of human action. We have argued why researchers should integrate insights about the representation of information with the mechanistic explanation of a task, considering representation as another causal factor. Furthermore, we have expanded the theory of mechanistic explanation by presenting four potential modifications of an explanatory mechanism, to be used for the explanation of development and learning. Applying this explanatory method, we have proposed to

explain the stable results of such dynamic processes as effects of kludge formation within a mechanism responsible for a given task. In addition, it was emphasized that both kludges and the associated representations can become generatively entrenched in the mechanism, giving rise to and taking part in subsequent developments, thus engendering a snow-ball effect. These insights regarding methodology and development were then applied to and helpful for explaining how actions are dependent upon different kinds, or levels, of intentions. We have introduced the concept of ‘sculpting the space of actions’, which enabled a comprehensive account of intentional action and the effects of expertise on it. Finally, we argued why the narrative configuration of action should be added to the intentional cascade, which also contributes to an expert’s sculpted space of actions.

It is not uncommon to complain that philosophers are better at raising new questions than at answering current ones. We hope to have shown in a modest sense that these in fact cohere intimately: by asking attention for issues or relations that have been neglected somewhat, existing problems often appear in a new light. Philosophical contributions can in that sense make useful contributions to the complex interdisciplinary investigation of human action, inviting as it does adequate questions and answers from as fields as diverse as philosophy, cognitive neuroscience, social science, robotics, computational and animal studies, and more. However, scientific projects rarely lead to the development of genuine interdisciplinary and comprehensive accounts but focus instead on further clarification of more specific features or elements. This can easily lead to rash and simplified accounts, which certainly has happened with respect to human action. The paradox of expert action, with which we started this section, is a point in case, as it depends in part on a misunderstanding of what lies behind automatic expert performance. Similar examples can be found in the debates about free will, a topic that we had to leave to another time. Yet the observation of expert action as a result of an agent’s long-term deliberately sculpting the space of his actions should make us pause about the rejection of the importance of free will for human action in general. In contrast to many who decry human intentional and voluntary action as being inexistent, impossible, outclassed or otherwise absent, this dissertation can also be read as an argument that intentional action is in fact possible, yet reliant on mechanisms and intentions that are more complex than often assumed. Indeed, our argument may even be taken as supporting the importance of not only musical but also moral education and practice: the admiration we feel for an expert opera singer or a moral hero is more than justified and should inspire us to likewise sculpt our space of actions.

**Figure I.** Representation of an explanatory mechanism responsible for  $\Phi$ -ing and some of its possible modifications, as discussed in Part I.

**Page 371-372.**

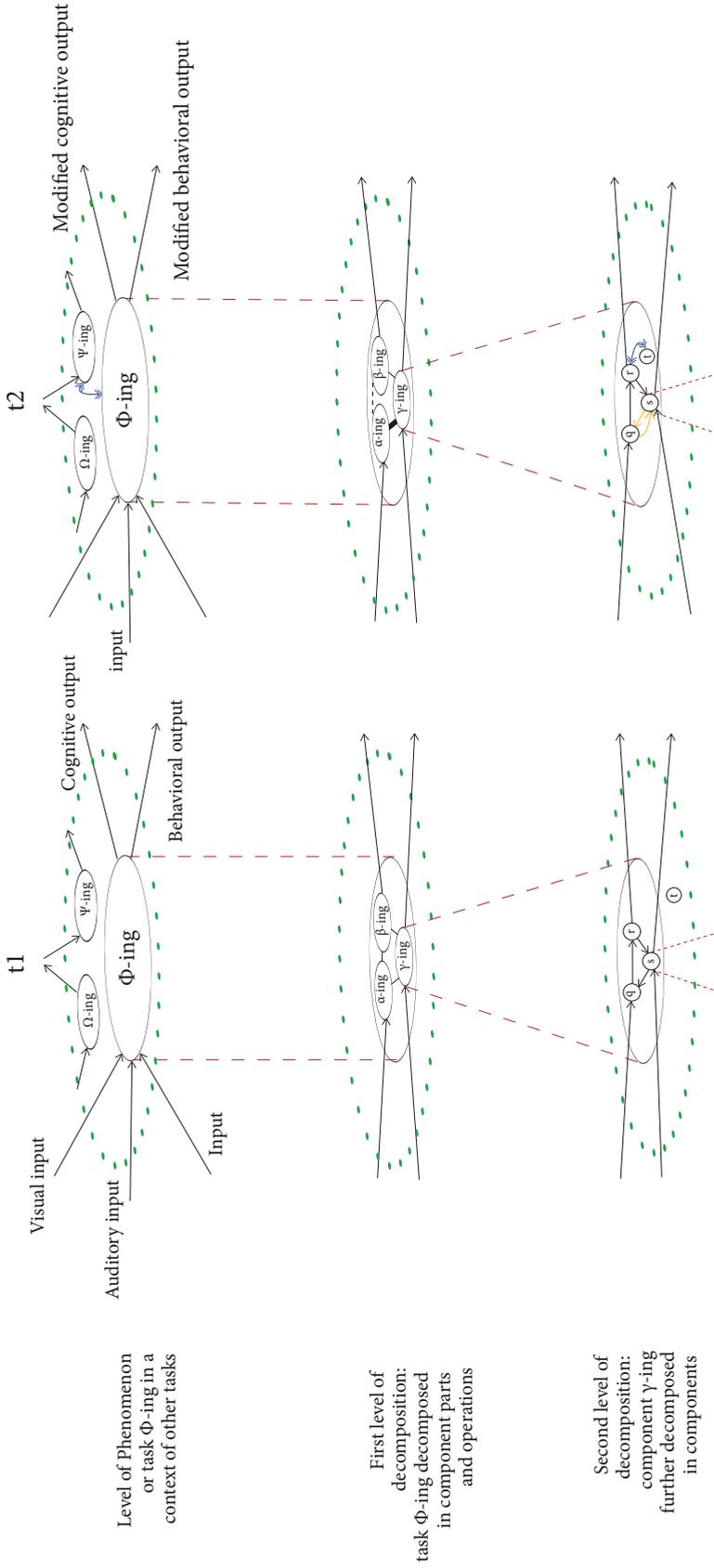
**Figure II.** Representation of an explanatory mechanism involved in development or learning as accounted for by neuroconstructivist theories, as discussed in Part II.

**Page 373-374.**

**Figure III.** Representation of explanatory mechanisms responsible for  $\Phi$ -ing in a novice and an expert in a particular situation with their distinct sculpted spaces of actions, as discussed in Part III.

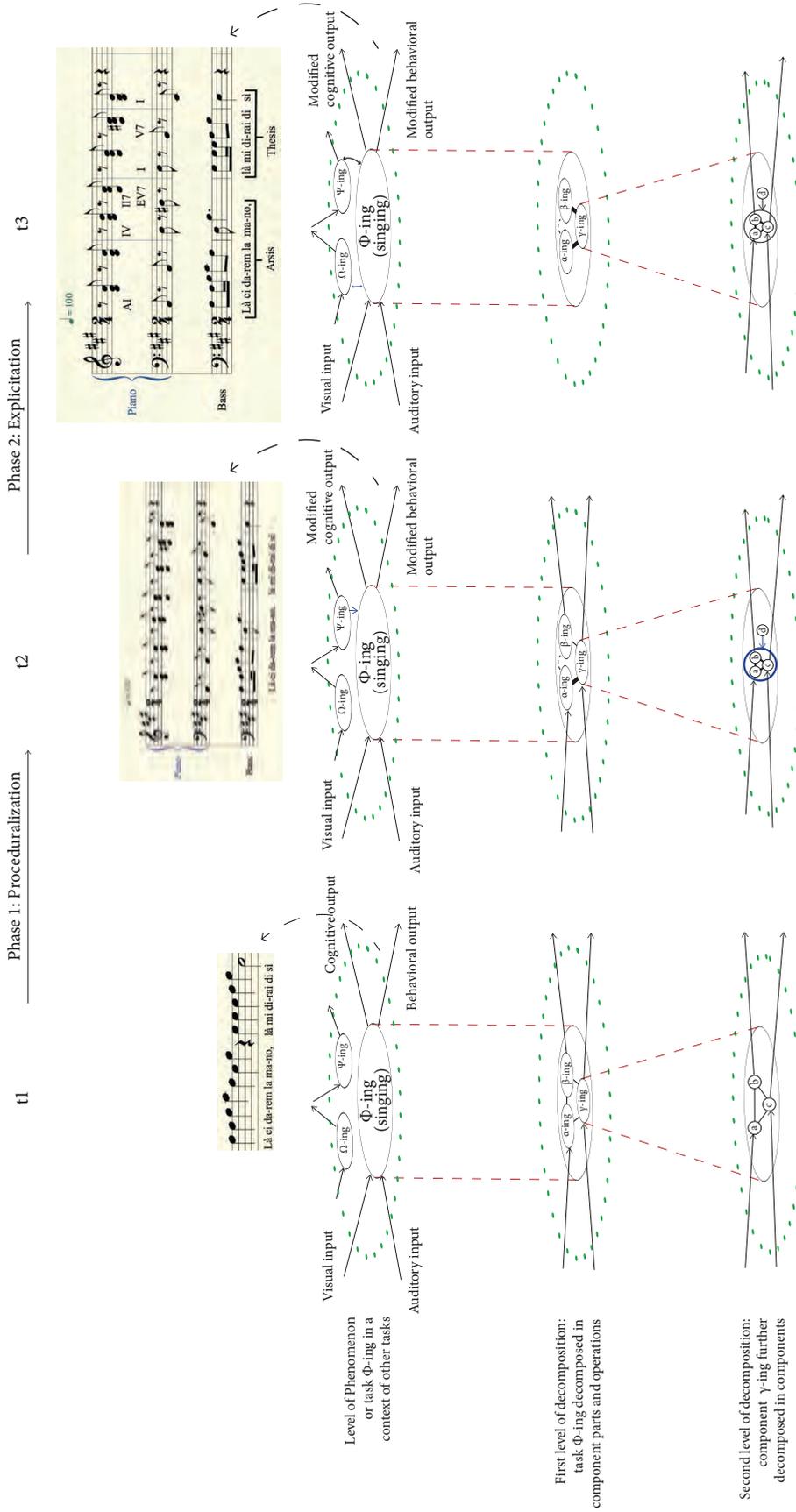
**Page 375-376.**

Figure I. Representation of an explanatory mechanism responsible for  $\Phi$ -ing and some of its possible modifications, as discussed in Part I



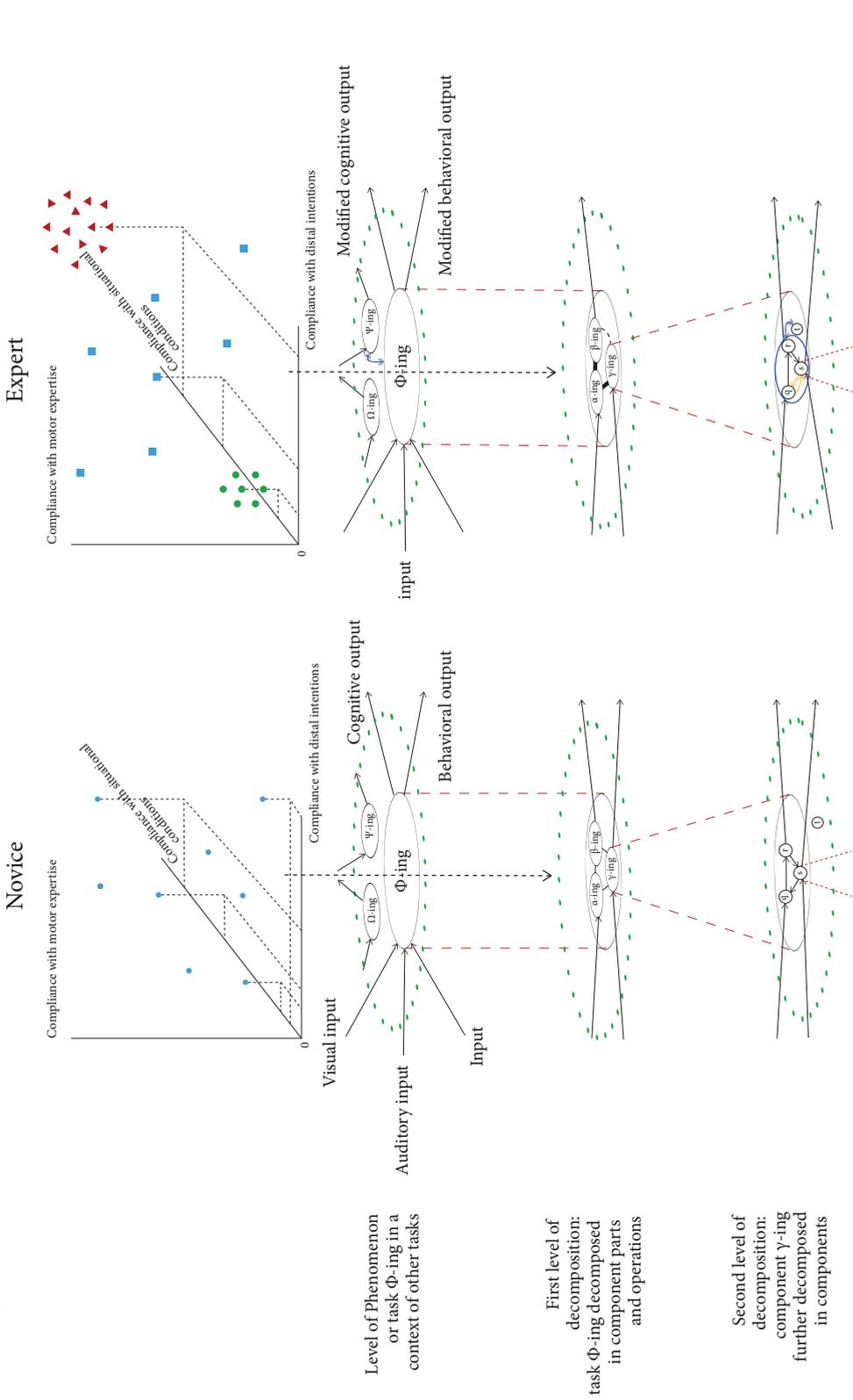
Simplified representations of the explanatory mechanism responsible for the task - or explanandum phenomenon -  $\Phi$ -ing at two different moments in time,  $t_1$  and  $t_2$ . Between those moments several kinds of mechanism modification have occurred at different levels, as was discussed in chapter I.5. Mechanism modifications represented here are: increased (bold) and decreased (striped) interactions between components; a new feedback relation (yellow) between  $\psi$  and  $\tau$ ; new influence (blue) of component  $\tau$  on component  $\psi$ ; new interaction (blue) between  $\psi$ -ing and  $\Phi$ -ing, yielding an - indirect - influence of new input to  $\Phi$ -ing. These mechanism modifications are together responsible for modifications of the cognitive and behavioral outputs of  $\Phi$ -ing. Note that vertical red dotted lines represent constitutive relations between mechanism levels and that green dotted circles represent the context within which components figure.

Figure II. Representation of an explanatory mechanism involved in development or learning as accounted for by neuroconstructivist theories, as discussed in Part II



Simplified representations of the explanatory mechanism responsible for the task  $\Phi$ -ing, in this case, singing. According to the neuroconstructivist theories of development and learning discussed in chapter II.2, two (partly overlapping) phases of learning can be distinguished. Mechanism modifications are involved in this (see figure I). The first phase of proceduralization is characterized by improved performance of  $\Phi$ -ing, enabled by the modularization (blue circle) of some mechanism (sub-)components which enables their faster and stable processing, enabling increased (bold) interactions between components at another level. This also facilitates additional interaction (blue) between tasks  $\Phi$ -ing and  $\gamma$ -ing (e.g. acting). The second phase of explication is then characterized by increasing explicit control of  $\Phi$ -ing, again facilitated by additional interactions (blue) between tasks. Due to representational re-descriptions, multiple representations involved in the task (curved dotted arrows) are available to the agent, leading from an implicit and simple representation to more complex, hierarchically structured and explicit representations of the music piece. Note that vertical red dotted lines represent constitutive relations between mechanism levels and that green dotted circles represent the context within which components figure.

Figure III. Representation of explanatory mechanisms responsible for  $\Phi$ -ing in a novice and an expert in a particular situation with their distinct sculpted spaces of actions, as discussed in Part III



Simplified representation of the explanatory mechanisms responsible for the task - or explanandum phenomenon -  $\Phi$ -ing in a novice and an expert respectively, in this case the task of determining an action in a particular situation, as was discussed in chapter III.4. With his increased expertise, the expert's responsible mechanism has become modified (see figures II and III). Corresponding with that process, the space of actions from which an action will now be determined by the interacting (sub-)components of  $\Phi$ -ing has become sculpted in the expert, which is not the case for the novice's space of actions. Red dots represent preferred action options that maximally comply with his motor expertise (horizontal axis), with his distal intentions (vertical axis) and with the situational conditions (diagonal axis). Green dots represent suppressed action options that comply only minimally. Blue dots refer to action actions that are neither preferred nor suppressed. The novice's space of actions contains only a few of those blue dots. Note that vertical red dotted lines represent the context within which components figure.