



UvA-DARE (Digital Academic Repository)

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

Keestra, M.

[Link to publication](#)

Citation for published version (APA):

Keestra, M. (2014). Sculpting the space of actions: explaining human action by integrating intentions and mechanisms. Amsterdam: Institute for Logic, Language and Computation.

General rights

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

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

1 INTRODUCTION: MULTIPLE MECHANISMS YET STABLE PATTERNS*

The riddle which has been mentioned at the beginning of this dissertation and has guided us through the previous Parts will be put more in central focus in this Part. The riddle amounts to a paradox: to which actor do we ascribe more intentional control of his actions – to the expert who performs a complex action without the continuous, conscious selection and control of his actions, or to the novice who is almost incapable of performing that action as he has to continuously select and control all his movements and vocal sounds? On the face of it, one would perhaps ascribe more intentional control to the latter, but a second look offers good reasons for preferring to ascribe maximal intentional control to the expert. Nonetheless, perhaps some qualifying statements need to be added. But let us build up the question first.

The first steps of our argument did concern the kind of explanations that are available for cognitive functions, as they are underpinned by neural processes, influenced from without and within, and change over time due to development and learning. After considering different types of explanation, we argued that mechanistic explanation seems best capable of accommodating these properties of cognitive functions. Not only can we apply the three different theoretical perspectives articulated by Marr to all components that figure in such explanatory mechanisms, these are optimally prepared for accounting for the dynamical processes that cognitive functions are involved in. Learning to sing can be mechanistically explained as the recruitment of an additional component of tone control into the mechanism responsible for speech; steadfast voice control is then depending on the stability of the network that constitutes the mechanism; change in the mechanism's organization occurs when growing expertise is associated with automatization and the corresponding decrease in recruited network components; such a decrease does allow novel influences on the mechanism's activities, for instance when a singer's increased expertise enables him a flexible responsiveness to perceived orchestral sounds.

The second Part offered insights in how a single function can be performed by multiple mechanisms. This is the more so, as mechanisms – complex and dynamic as they are – will always develop patterns of stability, partly as a result of a process of so-called kludge formation. Whether it is child development of voice control, or the automatization of certain associated patterns of stereotypical behavior, or the establishment of simulators that help to smoothly and comprehensively interpret and play a scene from an opera, or the seamless integration of external tools and objects

* 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. Fig. III is particularly relevant as a representation of the main contents of section III.4.

in an actor's performance: the formation of kludges in the responsible mechanisms facilitate such developments. Due to the formation of these kludges, several additional characteristics could be added to the dynamical mechanisms involved in complex behavior. A few are particularly interesting for the topics of this Part.

First, even though we can observe the emergence of a kludge in the changing properties of the performance of a cognitive or behavioral function - particularly as the function is performed more coherently and consistently - we cannot derive from these properties the representations or information processes involved. Neither can we derive from the observable properties the neural implementation of the responsible kludge. Indeed, processes like development or automatization, described in the previous Part, may appear similar with respect to several properties even though the underlying representations may differ from case to case, as may their neural implementation.

Second, kludges emerge in the mechanisms not from scratch but usually by recruiting or re-using components that were already in place - corresponding to the modes of mechanism modification that were mentioned in Part I. Similarly, these kludges can themselves be involved in further developments and present building blocks for future changes in the mechanism. This is one of many phenomena that support the assumption that hierarchical structure is prevalent in many of such mechanisms and in the actions they perform. Indeed, it was mentioned several times that, generally, we can observe that early developments become so much involved in multiple later ones, that the early components become ever more generatively entrenched in the mechanism. For example, once a child's language production has become stabilized, the underlying mechanism will be entrenched ever more and can not easily be changed in an essential form. Consequently, a change in his language production will have a much larger effect on a person's overall performance than a change in a more superficial and recently developed capability.

A third and final - of the seven - characteristic to be mentioned here is the involvement of external information and even objects in kludge formation. Kludges are not only established due to the internal repetition of certain neural network activities, but also under the influence of external information in most cases. Whether it is the implication of culturally specific tone systems in early learning of vocal control, or the association of the manipulation of a sword or cross with difficult melodic lines, external information leaves eventually some trace on the kludge that is established during the learning process. As a result of this, we should recognize how an explanatory mechanism is partly characterizable in terms of the external information involved in its development.

In sum, until now we have more generally discussed the theoretical and empirical arguments that support the notion of cognition and behavior being produced by modifiable mechanisms that can establish kludges with the involvement of previously developed (or evolved) components and of external information and objects. Indeed, we have argued that the complex performances that an opera singer must give are made possible by this mechanism modifiability. Integrating speech and singing with acting and responding to other protagonists and the accompanying music, while flexibly adopting the interpretation of the director, an opera singer can do this only provided his brain is capable of developing mechanisms that include these kludges. So far, we have barely touched upon the question what role intentions play in all of this. On the contrary, it may have seemed that intention did not play a role earlier as we were mainly interested in phenomena like proceduralization and automatization.

Intentions will be put more central in this Part. To that end, we will discuss a differentiation between different types of intentions and scrutinize their interrelations. This discussion will be nourished both by philosophical analyses of action and the intentions that partly determine it and by empirical investigations of the processes involved. As we will find, the philosophical analyses are not only directed at the level of an agent's explicit formulation of intentions to act, but some also aim to clarify the contribution of component processes to his action – component processes that in themselves resist verbal articulation, like the motor intentions or the representations of motor movements (Pacherie 2008). We will apply and discuss the framework offered by Pacherie, which describes a 'cascade of intentions' as it offers an integration of three different types of intentions that together allows an analysis and explanation of human agency (Pacherie 2008). As these different types of intentions can all play some role in the processes that – sometimes after extended periods of time – lead to an action, the framework facilitates systematic discussion of these roles with reference to both philosophical and empirical insights pertaining to them.

Indeed, given the results of Part I we expect to find that a mechanistic explanation of action will indeed refer to several component processes that in an organized fashion, including interactions between these component processes, produce an action. Following up to the issues discussed in Part II, we will also investigate whether this framework allows room for the modifications that we found to take place during development and learning. Kludge formation, we did find, can have profound impact upon the mechanisms underlying the performances of an agent. Regarding the present context this raises the question whether such kludge formation has an impact on how an agent determines and configures his actions, affecting their complexity and temporal extendedness. What would it mean for this cascade of intentions when

specific actions have become automatized? Once his expertise has resulted in a large set of actions or a domain of action that have been acquired, practiced, and modified, will this have an impact on the intentions that he forms? Are these intentions thwarted by the prominence of these actions, which are often performed automatically, or is the converse the case: an experienced agent is much better equipped for intentional actions even if these do not always require his conscious control.

These and other questions will occupy us for the rest of this final part of our dissertation. We will follow the structure offered by the intentional cascade and amend it in certain respects. However, next to the distinctions it makes between distal, proximal and motor intentions, we will first explicitly determine the notion of a 'sculpting' process which leads to an agent's 'sculpted space of actions' into this framework. Furthermore, given our interest in the mechanistic explanatory approach, the question is whether the discrete distinctions suggested by categories of intentional action are correlated with different cognitive processes and perhaps also with corresponding different forms of neural activations? Moreover, we will consider whether there is a correspondence between the hierarchies that structure both the intentional cascade and the mechanism underlying it. This is the more interesting, as one could suggest that actions performed by an agent are different from a novice's regarding the kind of intentional control to which they are subjected.

Let us now first clarify what we mean with a 'sculpted space of actions' before presenting the intentional cascade at more length. Together these two discussions will allow us to formulate some features of the framework that we consider necessary for the explanation of an agent's increasingly complex performances.

1.1 'Sculpting the space of actions' – an important ingredient for the explanation of expert action

Human action allows no simple explanation, as a causal pluralism is involved in its production and we argued in Part I that a corresponding theoretical pluralism is required for its explanation. We rejected a Socratic approach to intentional voluntary action, that holds that it is rationally deduced from an absolute moral principle. Instead, we embraced the Aristotelian account with its causal pluralism, according to which, for example, moral principles can somehow become internalized in an agent's habits and dispositions, implying that these principles can exert their influence via more than just a single representational format and a single process. Apparently, moral deliberation can yield results that are somehow accessible and useable by psychological processes as different from such deliberation as habits and dispositions. In sum, different as these processes and the contents involved may be, they interact and influence each other

and together result in an action. Modern studies have confirmed the causal pluralism behind action. Neurophysiological processes, affective motivations, memories and expectations, rational deliberations, environmental conditions, social influences and many other factors interact with each other during the performance of an action. In addition, development and learning create differences between novice and expert action in many ways, as we learned in Part II. Representations and mechanisms change and influence the properties of an agent's cognitive and behavioral responses, adding complexity to the pluralism. Yet somehow all of these factors converge and together produce, or result in, an action. So how should we conceive of this complex and dynamic process, given the various responsible component mechanisms involved in it and the various types of informations and their transformations that are being considered by the agent?

We propose to view this problem of determining an action by an agent as a search for a suitable candidate action in a multidimensional space – a space of action options. This action space is influenced by a multitude of factors, both dependent upon the agent himself and upon external factors. The action space's shape will be modified in a relatively stable way due to development and learning, yet will remain adaptive as it also responds to ongoing internal and external conditions.

To offer a first explanation of this view, let us consider a similar framework applied to a language processing task. Take, for example, the fact that we are capable of speaking fluently and use thousands of words while doing so. Now finding a word to begin or continue a sentence can also be considered as a search problem: a problem to find one or more suitable options in a large space of options. Particularly for those with expertise in a language, there are usually many alternatives for each word, even more when they have mastered several languages. These alternatives are not identical, though, differing in terms of semantics, grammatical and syntactical properties, idiomatic meanings, associations, and so on. Each of those factors can function as a constraint on the space of options available for an appropriate word that is to be expressed, constraining the search problem somewhat and alleviating the task accordingly. For example, when we write English sentences, the space with suitable options is restricted, as Dutch, German, Greek and other words are excluded from it. At times, a multilingual speaker will inadvertently insert a word from another language into his English speech – the surprise and annoyance about this signals the fact that it is quite exceptional. Similarly, once we have chosen to use the *pluralis modestiae* or *pluralis auctoris*, words like “I” and “me” and the like are excluded – though these may appear in quotes. In a dissertation, expressions like “I just believe that...” or “It is stupid not to understand” are to be avoided even more than grammatically incorrect sentences. In short, there are many

constraints at work at several levels of specificity and different in kind, that can limit the search problem somewhat – if, that is, the author’s grasp of the language and the conventions of the trade is appropriate.

Studies with word generation tasks by Chris Frith and others confirm the notion that finding a word in a space of options is influenced by constraints upon that space. Completing a sentence or filling in a blank in a sentence has been shown to involve various component tasks like generating words, selecting from a particular set of words, checking words for different sorts of appropriateness, inhibiting inappropriate options, and so on. Associated with these tasks, multiple component mechanisms have been identified, along with multiple representations and transformation of contents at several levels of specificity. Frith explains the fact that dorsolateral prefrontal cortex (DLPFC) activation increases in cases where there are only few constraints on the space of options provided by a particular sentence. Here, the subject somehow has to determine by himself (or herself) a selection of options and pick a final answer, and DLPFC appears to be involved. Frith describes this as: “the “sculpting” of the response space normally achieved by external context that has to be self-generated” (Frith 2000 560).

‘Sculpting the response space’ is understood as a dynamic process that is determined by factors that can be internal or external to the subject, stable or dynamic. Internal factors that influence this sculpting process are the language expertise of the subject but also cognitive and neural factors like his memory and the current stress hormone levels. External factors matter, as when the subject is confronted with easy or difficult task sheets, but the treatment by the research assistant also influences the process via the stress responses that it provokes. Some of these factors are relatively stable, while others dynamically influence the process, for example a given situation or even a particularly shocking word in a particular sentence. Each factor refers to a different (component) mechanism that will involve a specific representation and transformation of information. Nonetheless, as finally a particular response must be given, all factors must somehow converge in the process of determining a single option from the response space – each in its own way.

Such a response selection from a large space of options is relevant in many domains other than language. Indeed, this process of ‘sculpting the response space’ is held by Frith and others to obtain in different modalities and task domains, including action selection (Frith 2000).²⁷⁹ They have devoted some research to further determining the set of component tasks and elucidating the impact of particular constraints on this modification process (Fletcher, Shallice et al. 2000). For example, a study with different representations of information and their transformation has shown how these

influence the process, too. Rule-based selection turns out to be important, for example, recruiting particularly DLPFC activation, while other component tasks involve other forms of association and are carried out by other neural activities (Nathaniel-James and Frith 2002).²⁸⁰ Indeed, studies like these inform us about the mechanisms and representations involved in particular instances of sculpting a response space, shedding light on the specific properties of the space itself: its dimensions, its flexibility, its structure, and so on.

As attractive as this framework of a search problem being influenced by a sculpting process seems to us, it has been applied only to a limited extent in the domain of human cognition and action but perhaps more in the domain of AI and robotics.²⁸¹ Let us consider a few prominent examples of such applications before arriving at our notion of ‘sculpting the space of actions’. The explanation of human color perception may be the domain in which a spatial or geometrical framework is most widely used. For example, colors have been taken to be points in a multidimensional space determined by their various properties, like value, hue and chroma (Munsell 1912). Meanwhile, certain properties of human color perception are explained with reference to a spatial color representation in the form of a spindle.²⁸² In this way, a range of phenomena can be explained, like the perception of after-images, the effects of contrast colors and the differences between languages in how they carve up this space with their respective color vocabularies (Regier, Kay et al. 2007).

²⁶⁷ Another experiment in which subjects were asked to engage in ‘willed action’ by lifting at random their fingers showed DLPFC activation patterns comparable to those that occurred in word generation tasks. This concurred with observations in other experiments that required subjects to act or move (Nathaniel-James and Frith 2002). Indeed, there is growing consensus with regard to the overlap in neural activations for language processing and action processing, see for example (Grèzes and Decety 2001 ; Pulvermüller 2012; Pulvermüller, Hauk et al. 2005 ; Raposo, Moss et al. 2009 ; Taylor and Zwaan 2009 ; Willems 2009). This suggests that the representation of the content of speech and action do overlap to a large extent. An implication is that such representations are available for several processes, offering a crucial role for simulation as a prevalent form of computation (Barsalou 1999c ; Jeannerod 2006).

²⁶⁸ Interestingly, in conditions when the response space is less constrained – by contextual clues, for example –, subjects show more DLPFC activation and take longer to respond (Frith 2000).

²⁶⁹ Supporting the usefulness of multidimensional spaces as the representational format of information involved in cognition and action is its feasibility for the construction of simulation and robot models (Gärdenfors 2004b ; Gärdenfors and Williams 2003). In a robot, for example, decision making can take place at the level of conceptual spaces, where all relevant information is represented. Here, perceptual constraints determine a set of possible actions that are preselected on the basis of the explicit programmed instructions (Chella, Gaglio et al. 2001). Artificial decision procedures can be developed in this vein as well. For example, expert systems in clinical situations make heavy use of ‘relative magnitudes’ pertaining to specific dimensions and which stem from counts, measures, weights, and so on. Gärdenfors’ framework as a ‘meso level representation’ is considered helpful for such construction work (Aisbett and Gibbon 2001). In all cases, however, the number of dimensions and their relations have been decided upon by an engineer and have not evolved and developed naturally, nor are they allowed to be modified in unpredictable ways as is the case in human agents due to their development, learning and experience.

Another example concerns taste. It is argued that particular tastes are the result of specific activation patterns elicited by the four taste receptors and consequently occupy specific areas in a multidimensional ‘taste space’: “[i]n this way are the brain’s representations of the various possible tastes arranged in a systematic “space” of similarities and differences” (Churchland 1995).²⁸³ Although these taste perceptors are activated by different electrochemical processes, their activations are represented in a common space and arranged together in such a way that they give rise to our recognition of thousands of different tastes.

In a similar vein, Churchland has applied this framework to explain how an agent determines his action in a world full of objects. Instead of separating the processes and representations involved in his perception of his environment, his decisions to move and the performance of the movement, Churchland argues that these processes should be taken as employing a shared multidimensional space. Evidence for this comes from studies demonstrating the interactions and interferences between those processes, among others. This implies that the spatial representation of object location and the spatial representation of bodily movement are somehow integrated in the same representation space instead of employing two separate spaces. If the latter were the case, it would require a complex translation process of his spatial movements for placing them in a space that represents the environment, if the agent wants to avoid bumping continuously into external objects. Indeed, the integration of both representations into a single space would facilitate his sensorimotor coordination and thus enable the agent to: “assume[...] a position in its “motor space” that corresponds to the position of an object in its “sensory space” (Churchland 1995).²⁸⁴ Evidence suggests that this integration is indeed the case.

²⁷⁰ It is debated whether color vision or taste are more dynamic processes than is often thought, affecting their representations too. Investigation of neural firing rates related to taste perception in rats under different conditions, demonstrates that taste perception is a highly dynamic process, modulated by other cognitive processes or states of the animal. As a result, hedonic impact and incentive salience of a taste are variably modulated (Tindell, Smith et al. 2009). Modelling such changes via adjustments of a state space would require continuous modifications of the geometry and topology of that space, leaving behind some of the attractive simplicity of such representation.

²⁷¹ The spatial arrangement of taste representations is also used by Churchland to distinguish between tastes that are prototypical for a particular fruit, for example, hyperbolic divergences of these, and so on. In addition, with such a spatial format of representation, relying as it does upon the number of dimensions combined with the levels of discrimination within each dimension, we can also determine the size of a state space pertaining to a particular cognitive function and compare it with a space of other functions or with a similar space in other animals. The state space of smell, for example, is much larger and contains much more levels of discrimination in dogs than in humans (Churchland 1995).

²⁷² Compare this issue with the theory that sensorimotor coordination is enabled by ‘common event codes’, that is, by codes of features of perception and action plans stored in a common representational medium (Hommel, Musseler et al. 2001). This theory is modest in its domain of application and specific in that event codes are presumably shared by both tasks and how they are implemented.

It is important to realize that such a multidimensional space is employed here as an explanatory tool. It offers a plausible representation of how the results of different cognitive processes appear to be related to each other. In other words, it is a second order representation, a representation of the results of multiple cognitive processes. Such a second order representation is different from the two representational formats that are prevalent in cognitive neuroscientific explanations: the representational format of symbols and propositions and the representational format of connections among neurons.²⁸⁵ An important characteristic of the multidimensional, spatial representation is that a fundamental role is played by similarity-dissimilarity relations. In this way, we can explain how color perception appears to employ a continuous color-space, while this space simultaneously appears to be carved up into separate sub-spaces with verbal categorization.²⁸⁶ Such an explanation of color concepts within the same framework as color percepts is parsimonious, indeed (Gärdenfors 2004b 2). This framework has been used to explain other features of cognitive processes as well.

Important for our purposes is whether dynamic factors like someone's expertise or an environmental condition are allowed in this framework, changing for example the contents and shape of his or her representational space. Gärdenfors indeed argues that learning or development (and even evolution)²⁸⁷ can be explained in terms of the change of contents or structure of a relevant space.²⁸⁸ Such changes in a person's representational space for a particular domain can be stable but there are also dynamic

²⁷³ Gärdenfors concurs with the critique of the sentential conception and the defense of a spatial account of information representation given by both Churchlands (Gärdenfors 1996).

²⁷⁴ A category is a particular region of a conceptual space that has been carved up. Even though the conceptual space is itself continuous, it can be carved up into regions with sharp boundaries. For example, even though the color space is continuous, color terms suggest sharp boundaries between colors (Gärdenfors 2004b). Probably due to visual physiology there is large agreement between languages in how their color terms refer to regions in the color space (Kay, Berlin et al. 1991). Nonetheless, there are differences between separate languages in the number of color terms they use or the boundaries drawn between color terms. Interestingly, experiments with English and Korean subjects shows that 'categorical perception' occurs as subjects' perceptual discrimination corresponds with the category boundaries of their language, suggesting an important role for 'categorical perception' (Roberson, Hanley et al. 2009).

²⁷⁵ Evolutionary processes have also effectively contributed to the determination of spaces for several functions. Gärdenfors argues that one could easily reformulate Marr and Nishihara's explanation of visual object shape recognition in terms of the employment of a 'shape space'. According to them, evolution seems to have resulted in a rather simple process depending upon the fact that biological objects tend to have a form based upon generalized cones. Consequently, for visual recognition only a limited set of dimensions needs to be processed, like the size of the cones that are connected to each other, their orientation, the components' axes and their reciprocal configuration (Marr and Nishihara 1978). This approach therefore allows the representation of biological objects within a multidimensional space of options.

²⁷⁶ The author defends the representation of Piaget's findings of how children learn to differentiate between height and volume when perceiving filled glasses according to his geometrical account of cognitive processes (Gärdenfors 2004b). To be sure, this development can also be represented differently, for example as an event along the lines of catastrophe theory (Molenaar 2001).

changes of his or her representational space that obtain as a result of environmental or internal processes. For example, a process like the direction of attention to a particular property can be considered as affecting the agent's representational space. When he is attending specifically to an object's weight he can make finer discriminations, which can be represented as 'stretching' the distances along that dimension. Conversely, neglecting such a property amounts to 'shrinking' these distances (Gärdenfors 2004b 20).²⁸⁹ Many more internal and external can contribute to such changes of an agent's representational space, affecting among others the response space that his cognitive or behavioral response depends on.

The discrimination of actions has been described in terms of such changes of a representational space, brought about by a possible neural network which learns to distinguish two dimensions. More specifically, within this spatial framework it is explained how the network might represent and discriminate between moral dimensions of an action. By judging the similarity and dissimilarity between actions along some relevant moral dimensions, the network would place actions like assisting, murdering, lying and self-sacrifice at various locations in an action space.

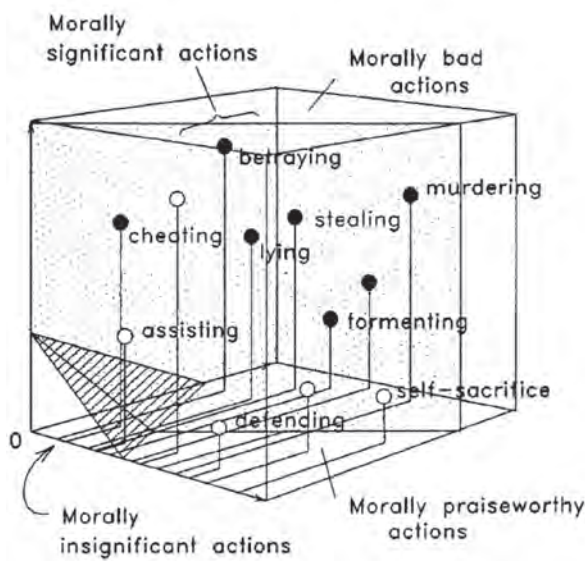


Figure 2. A (conjectural) activation space for moral discrimination

Reprinted from (Churchland 1998) with permission from the publisher.

²⁸⁹ In Churchland's terms, such refined discriminatory ability would amount to an increase in the levels of discrimination with respect to a certain state space or one of its axes (Churchland 1995)

Paul Churchland and others argue that an agent does not usually determine his actions via deduction from abstract moral principles but by employing such a framework of action representations in an action space. That is not to deny that irrespective of the spatial contiguity of this space, moral concepts and judgments tend to carve it up rather strictly, as do color concepts with the multidimensional color space (Churchland 1998; Casebeer and Churchland 2003). In figure 2 above, this is represented by the vertical pane that separates praiseworthy from bad actions and by the diagonal pane that leaves a corner for morally insignificant actions.

What is not visible in this framework for the spatial representation of moral actions is whether its representations are flexible as a function of the conditions under which it is employed. For example, we would maintain that for most agents, no action is morally insignificant under all circumstances – and perhaps vice versa. Moreover, when an agent is about to determine an action, the dimensions of his space of actions will flexibly respond to internal and external conditions like his emotional state or the risks provided by the environment. These conditions would have an impact on the dimensions and the structure of the spatial representation and with that, also on the placement of the actions in it. It may even be the case that the conceptual distinctions he usually makes – represented as panes in the figure – will shift or that an individual action will change sides.

Let us now apply what we have learned from this short discussion to the notions that we are introducing in this dissertation: sculpting the space of actions, and a sculpted space of actions. Whereas Frith's framework of 'sculpting the response space' primarily referred to an ongoing dynamic process of determining an appropriate answer for a given problem, the other spatial representations rather contained stable representations of a domain like color or moral action as used by an agent. Frith's framework implied that internal and external constraints help to constrain the response space in a particular case and facilitate a final response choice. The other frameworks emphasized how the stable representation of a domain can change due to development and learning in terms of its size, its dimensions, its structure and so on. Our framework, finally, has the ambition to combine these two, realizing that dynamic and stable properties of a sculpted space can and should not be separated but integrated.

Indeed, we noted that according to Frith, the process of 'sculpting the response space' was also constrained by the relatively stable properties of the space due to the agent's expertise. These stable properties are themselves the result of a long-term sculpting process, since development, learning and practice will affect an agent's representational space pertaining to a particular domain. The more this has resulted in

a stably 'sculpted space,' the easier it usually should be for an agent to find an appropriate response.²⁷⁸ In other words, finding an answer to a particular question amounts to dynamically applying *further* constraints to a representational (sub-)space that is itself already sculpted or constrained as a result of expertise. For example, the recognition of relevant constraints in a given situation depends upon the agent's mastery of the domain, such as by his knowing musical structures and rhythms.

Yet this interaction between the stable and the dynamic properties of a sculpting process is itself complex.²⁷⁹ Irrespective of the agent's mastery, this recognition of relevant constraints can easily be impeded by other cognitive and neural processes that simultaneously take place. Attention, for example, was already mentioned to potentially affect a dimension of the representational space employed by the agent, stretching or shrinking it. Stress was mentioned as well, as a dynamic factor that can influence several processes involved in the sculpting process. However, in the case of an expert, we expect him to be less vulnerable to such disturbing factors and to perform more reliably and appropriately than a novice, attacking the correct note at the appropriate time, for example. In other words, the stable space that he has established over a long time of sculpting has such properties that it allows him to determine an appropriate answer even in cases where he is affected by distraction or stress. So, in an expert singer's sculpted space of actions the tritone or *diabolus in musica* occupies a very small place in a far corner of the space, making it unlikely for him to sing it even when he is tired. Moreover, his expertise should also imply that he realizes how such conditions dynamically affect the sculpting process, knowing strategies for countering them. Being tired, he pays extra attention to what the orchestra plays and

²⁸⁰ It might be maintained that there are cases in which an expert, having a sculpted response space, can come up with a response that is less adequate and innovative than a novice's. Such a response would be represented in a relatively small and peripheral subspace in the expert's action space. Such a case might indeed occur, yet it is likely to be a chance hit because what is characteristic of a novice is his limited expertise with regard to the constraints and properties of the domain. We concur with Boden, who argues that one should not consider such a novice's response to be genuinely creative (Boden 2004).

²⁸¹ These spatial accounts also emphasize the temporal dimension of cognitive processes, often by integrating some form of dynamical systems theory. This is offered as being superior with respect to its handling the temporal dimension in cognitive phenomena than traditional approaches like the traditional computational approach (Van Gelder and Port 1995). Such a dynamical systems approach can also be used without talk of representations, even though the latter is common in cognitive science. Instead of content being represented in a distinct way within a computational system, different vectors – for example, one for every taste receptor – determine the state that a particular system is in. In that case, the state of the dynamical system at any moment carries the necessary information without there being some sort of representation present in the system (van Gelder 1998). One could argue that this is not more than a semantic difference between the approaches, differing as they do in their definition of what a representation is. Furthermore, there are many cognitive functions in which correctness or incorrectness of the representations involved do matter. Dispensing with representations altogether would therefore not be advisable (Bechtel 1998).

to the preparation of his voice in order to sing the right note.

Finally, it is important to remark that this representation of actions in a multidimensional space of actions is not at odds with further references to other action representations. The multidimensional space allows us to add further properties or dimensions to the actions that are represented in it, as long as each dimension can be processed by a cognitive or behavioral function. For example, we will next discuss how an agent's actions are influenced by the different kinds of action intentions that he has, differing in their format of representation, among other things. Motor intentions, for example, are in the non-conceptual format of sensori-motor representations, whereas distal intentions are verbally formulated and in a conceptual format. Expertise with these different levels of intention regarding a particular action will affect the action's place in the agent's space of actions. For example, it may be that a novice has not yet practiced the sensori-motor representations that belong to a certain action, making it less likely that the action will be selected from his action space in an emergency. When an expert has extensive experience with that action and he is required to act, this action associated with sensori-motor representations will figure more prominently in his constrained space of actions. Expert action, then, is indeed dependent upon sculpting the space of actions along the lines described here. The sculpting process involves a combination of both expanding and constraining the space of actions. Important to repeat is the fact that this sculpting process is both a long-term and a short-term process: it contributes both to an agent's stably sculpted space and to the dynamic properties of his responding in a given situation. Without such sculpting the space of actions, it is difficult to see how expert action can be performed at all.

1.2 Determining an action via a cascade of intentions

An expert singer may have noted that his recent performance of Don Giovanni had unwillingly been affected by his solemn interpretation of Saint François not long ago. Intending to correct this flaw in his interpretation by making his Don Giovanni more boastful in general, he realizes that some solemnity may still work in the dialogue with Donna Elvira, who is naively trying to convert her unfaithful lover. Therefore, even after having formulated a general intention for a more boastful Don Giovanni, our singer needs to more specifically remain alert for situations in which this intention can prudently be acted out, while avoiding others. Needless to say, that such intentions and performances can only usefully be made by a singer who has such a mastery over his singing that he can switch timbres at wish from solemnity to boastfulness (and change his style acting correspondingly).²⁹² This description of a not uncommon form of self-correction and self-control involves three different types of intentions,

which we will introduce in this section. We can distinguish between these types of intentions in terms of their contents, their functional roles, their temporal trajectory, and so on. Important, however, is the fact that they are intimately related to each other, contributing to his coherent and consistent performances of different roles. Let us consider a model that aims to account for this impressive feat.

Integrating philosophical and empirical insights, a model has been developed to account for the hierarchical control of motor action via a 'cascade of intentions' that spans different levels of specificity of an intended action in (Pacherie 2006 ; Pacherie 2008).²⁹³ This model considers intentional action in its most comprehensive form as the result of a process that can be described as a process with three discernable and distinct phases, starting with a deliberate intention to realize a future goal and completing when particular muscular activities have realized that intention.²⁹⁴ Each phase has a different functional role, involving different formats of representation and transformation. These phases do not strictly succeed each other, with an intentional action possibly occurring without contributions of all phases of the complete cascade. In closing this short description of the framework, it should be noted that there are many different forms of interaction between the phases.

The model distinguishes between three different forms of intention, to wit: distal intention, proximal intention and motor intention. Distal and proximal intentions have been borrowed from several philosophical accounts of intentional action, while motor intentions were added to those on the basis of evidence from the cognitive

²⁸⁰ Research of expert performance in domains as far apart as sports, music and science has demonstrated that extensive periods of deliberate practice generate the necessary cognitive and physical adaptations for exceptional performance. Improved motor performance also requires enhanced cognitive representations and skills, contributing to its improved selection, guiding and correction of motor actions (Ericsson, Roring et al. 2007).

²⁸¹ It must not be left unnoticed that this three-level model of intentional action has been used particularly for the explanation of the phenomenology of agency. To that end, the model has been further equipped with feed-back and feed-forward relations between levels of intentions and comparators that serve to discern congruity or incongruity between intentions, motor movements, perceptions (Pacherie 2008). Apart from such specific use, however, it can still serve for the explanation of the 'generation and control of action' (Pacherie 2006). The model allows still further elaboration or expansion, for example with the integration of the What, When, Whether model of intentional action, as the authors of both models have hinted at in (Pacherie and Haggard 2010).

²⁸² There are several others models of intentional action available, of course. Depending on their explanatory or analytical focus, these models differ from each other. For example, the WWW model of intentional action does not so much focus upon the different forms of intentions involved, but rather on the different component decisions involved in intentional action, that is: the decisions about what, when, and whether to do (Brass and Haggard 2008). A different model of intentional action does also integrate cognitive – decision making - and motor processes like Pacherie's model does (Cisek 2006). These authors use the notion of 'representation' as a common denominator for all processes involved, eschewing the notion of 'intention'.

neuroscience of motor action control.²⁹⁵ Several philosophers of action have argued that it is important to distinguish between distal and proximal intentions since these play different roles in the various phases and features of actions.²⁹⁶ Most prominent, of course, is the difference in their being oriented towards future actions or being aimed at the realization of an action here and now in the case of proximal intentions. Relevant to note is that there is a reciprocal interdependency between the different temporal orientations of these two forms of intentions: if a proximal intention is carried out without any orientation upon the agent's distal intentions, the agent runs the risk of frustrating and even counteracting his own long-term intentions or might fail in the coordination with some other intentions of himself or interested other parties. On the other hand, realizing a distal intention requires the recognition of a suitable situation and anchoring the appropriate action in that particular situation, which is the role of a proximal intention. The performance of that appropriate action in a particular situation, which realizes the distal intention, finally relies upon specification of the necessary muscular movements that are captured by the motor intention.²⁹⁷ Given that such motor specifications, while necessary for flexible performance of intentional action, escape the kind of awareness and explicit control that can be applied to the other intentions, these differ sufficiently from those other two forms of intentions to merit separate mention.²⁹⁸

²⁸³ Pacherie (Pacherie 2008) makes reference to Searle's distinction between prior intentions and intention-in-action ((Searle 1983); see further below for its relevance in guided action), Bratman's distinction between future-directed and present-directed intentions (Bratman 1999a) and finally Mele's distinction between distal and proximal intention which was adopted by her (Mele 1992). Pacherie notes that it was mainly the absence of temporal connotations in the latter distinction that made her prefer it above the others. Our discussion, further below, will stress the importance of avoiding incoherence and inconsistency between actions, which is more at the center of Bratman's arguments.

²⁸⁴ Actually, Pacherie distinguishes even seven different functions of intentions: intentions can function both as prompters and as terminators of practical reasoning; they serve individual and social coordination; they can function as initiators of a performed action and serve to sustain an action until the end; performing meanwhile a guiding function while also assisting in monitoring the adequacy of the action's performance (Pacherie 2006).

²⁸⁵ Contrary to common use of what 'intentions' are taken to be, these motor intentions contain neither propositional content nor are agents usually aware of these in this framework (Pacherie 2008).

²⁸⁶ Within the modern Anglosaxon domain of philosophy of action, it may have been Frankfurt who was the first to point out the relevance of such automatic motor adjustments for intentional action. Moreover, he has also pointed out that the resulting 'purposiveness of our behavior' is not limited to human action as also spiders must be said to act intentionally, for example – albeit in a weaker sense than humans do (Frankfurt 1978). In continental phenomenology these topics had been debated much earlier, particularly by Husserl, Merleau-Ponty and some other phenomenological authors (Painter and Lotz 2007). This tradition has been inspired by Aristotle's philosophy of biology in which the animal's responsiveness to its environment figures prominently (Oele 2007). This is another example of a long forgotten lesson from Aristotle's biological works that has impeded philosophy, as it emphasizes continuity and gradual differences between different animal species and mankind instead of focusing mainly on distinctions and divisions.

Indeed, there is a dynamical interaction between all three forms of intentions. In fact, the interactions at stake can flow in two directions. To be expected is the top-down flow of control in which a distal intention refers to an agent's particular long-term goal, waiting for the appropriate situation to present itself for fulfilling such an intention by anchoring it in that situation by determining the corresponding proximal intention. Subsequent performance of the latter is dependent upon the motor intentions that specify the necessary movements. In addition, a bottom-up flow of control must be acknowledged, both on phenomenological and neuroscientific grounds. For example, the mere perception of an environmental object that can be involved in a motor action does provoke activation of the corresponding motor representations, without an agent's deliberate intention to act (Grèzes and Decety 2002). Indeed, the relative autonomy and independence of this lower level of action control is such that: "the affordances of an object or situation are automatically detected even in the absence of any intention to act" (Pacherie 2008 186).²⁹⁹ Subsequently, upon these motor representation activations, corresponding proximal intentions may arise, as when an agent may realize only his desire to quench his thirst when he finds himself reaching unwittingly to a perceived cup. Finally, an interaction between levels of intention may also occur in order to correct or interrupt an action, for example if motor movements must be adjusted or even interrupted due to a changing environment. More below in this Part we will further discuss such dynamical interactions.

Summing up the foregoing, we can refer to a 'cascade of intentions' that together comprise a hierarchical model of action control as is visible in Figure 3 below from (Pacherie 2008). In this figure, along the vertical axis, we see how P-roximal and M-otor intentions are subsumed under D-istal intentions. Horizontally, we can discern how it may take a while for an intention (dotted line) before it enters into the process of undergoing the necessary transformation via situational anchoring or parameter specification (downward along vertical lines). As a result, an overt movement occurs. Conversely, as mentioned above, a bottom-up form of control can happen when in the absence of proximal and distal intentions, for some other reason an overt movement is made, triggering situational guidance and control such that the associated proximal goal is reached. At times, however, this movement may be interrupted or inhibited when the movements take long enough for an agent to exert an additional form of (rational) guidance and control, particularly when he becomes

²⁹⁹ The clinical syndrome of utilization behavior suggests that indeed automatic detection of a potential opportunity for motor action can then lead to an action without the agent having a proximal intention for this action. In such cases, an agent may put up a pair of glasses even though he does not need a second pair upon his nose (Sumner and Husain 2008).

aware of its contradicting his distal intentions. As all these processes have their own temporal constraints, it is among other things dependent upon the tempo of the events whether the complete cascade of intentions can unfold, or not.

Having presented this model, it is relevant to underline again the use of putting these three types of intentions that have been distinguished philosophically into a single dynamical model, even though several differences between them have been noticed. Remember that in our Part I, we noticed that for the explanation of a particular cognitive or behavioral function we can develop an explanatory mechanism that captures the function at several levels of mechanism. Obviously, all intentions are the result of some cognitive processes going on in the brain and their realization does equally require an execution that at least involves some motor processes. Correspondingly, we can take the philosophical analysis as a heuristic and investigate whether its ingredients allow to be integrated as components in a more comprehensive explanatory mechanism. This will by no means be an easy task, as there are many

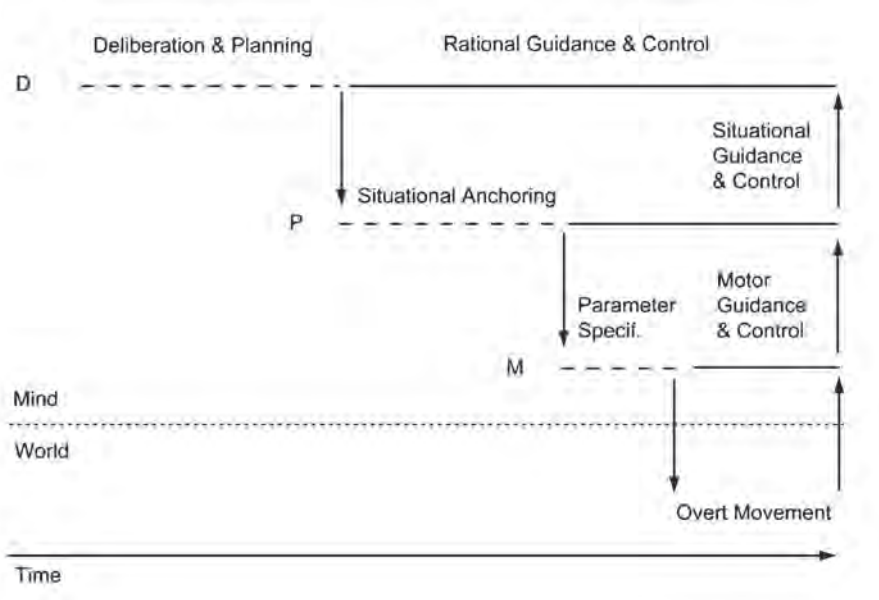


Figure 3. The intentional cascade of D(istal) intentions, P(roximal) intentions, and M(otor) intentions. Note that a horizontal *dotted* line refer to an existing intention still waiting for the phase of its further realization. A horizontal *continuous* line refers to the phase in which an intention is actually realized. Adapted from (Pacherie 2008 189) with permission from the publisher.

feedback and feedforward influences between levels involved, processes taking place at different time scales,³⁰⁰ and so on. Indeed, if an agent is to perform coherently and consistently, such influences and processes must be connected to each other. Therefore, instead of keeping a philosophical analysis of action intention separate from a cognitive scientific explanation of motor action, it is a challenge to see whether a mechanistic explanation allows us to integrate these. Indeed, such an integration would invite us to note that the ‘what’ or the goal of an action “can be specified at the three levels of M-intentions, P-intentions, and D-intentions” (Pacherie 2008 196). Similarly, the model would allow a specification of the ‘how’ or the means of an action at several levels of specificity, as it does of other action features.

With this model in place, we can analyze how the performance of an action is produced by a mechanism that consists of different interacting sub-processes, which can in turn be analyzed from different disciplinary perspectives – including a philosophical perspective. The model offers us also a framework to further explore some of the issues that we found to be relevant in explanation of cognitive functions. The first issue is the algorithmic theory of the nature of the representation of information. As the model suggests and Pacherie has also noted explicitly – see the quote in the preceding paragraph – , we can expect different representations at the three levels of intentions: verbal in the case of distal intentions and in the form of non-verbal motor representations in the case of motor intentions. It is particularly interesting to consider what representations are involved in the intermediate level of proximal intentions and we may expect an interesting confrontation between philosophical analysis and empirical insights in that context, given its position between those explicitly verbal distal and non-verbal motor intentions.³⁰¹ Since we’ve learnt in the first Part that it is impossible to directly derive this representational level from either the task level or the neural implementation level involved, investigating the actual form of representation is challenging.

The second issue which particularly interests us in light of our investigations in Part two, is how we can integrate in this model a central role for experience and action skills.

³⁰⁰ Indeed, apart from a control hierarchy that is responsible for the increasing specification of actions, it is important to acknowledge that actions also require a hierarchy of temporal extension as all actions are temporally extended. Both hierarchies do not necessarily overlap and require to some extent different neural and cognitive resources (Uithol, van Rooij et al. 2012).

³⁰¹ Pacherie refers to one of the two visual streams, to wit the ‘vision for action system’ in this context. This system allegedly produces motor representations in an appropriate format, usually involving an objects as an action goal, while taking into account several biomechanical constraints. The motor representations are then used in two different forms, as predictive or forward models and as inverse models (Pacherie 2006). The picture that emerges is still very much top down and still suggests a rather one-to-one correspondence between the distal and proximal intentions and their motor counterparts.

Given our findings regarding the gradual and dynamical processes of automatization and habituation of intentional actions and the corresponding differences between a novice and an expert singer, these processes could further demand elaboration of this model. As a result of the formation of several kludges and other changes through experience, we can observe in an expert how he has built up a space of actions that appears in many regards different from the space of actions the novice draws from. An important difference is that an expert can draw from this space without having to explicitly determine a particular action in detail, but switch completely his singing and acting from a solemn to a boastful Don Giovanni in an instant. Indeed, this space of actions can be ‘sculpted’ according to a particular dimension of the actions currently deemed important, to which then increased attention is directed. In such cases, a complex interaction takes place between an agent’s previously sculpted space of actions and his intentional cascade – a phenomenon that we will certainly consider more closely in this Part.

In sum, this model is interesting in that it suggests how we can combine both a philosophical account of different types of intentions with empirical insights in the determination or control of motor actions. At the same time, it invites us to further explore two issues that we earlier found to be relevant for explanation of cognitive functions: the representations involved in these and the role of experience and skill in their performance. In line with the latter issue, we will also be alert to find whether we can observe in the literature reports of the same benefits that we previously found to adhere to hierarchically structured complex mechanisms once they dynamically develop and change, for example by sculpting a space of actions. Observation of an expert opera singer suggests that he indeed has the advantage of such benefits, as does the audience that must not fear for being disturbed by any instability of his voice, by his incapability of harmonizing with the orchestra and other actors, nor by his forgetting his acting once the singing becomes difficult.

1.3 The cascade of intentions and a sculpted space of actions

What we are interested in this Part is the complex process that leads up to an action which includes the involvement of different representations of action in and the role of a sculpted space of actions therein – which probably also adds to the multiplicity of action representations involved. At first sight, a philosophical analytical approach is most appropriate for analyzing the explicit and verbal or symbolic representation of action, as it is found at the level of distal intentions in particular. However, the model of the intentional cascade discussed in the previous section has been developed by combining further philosophical insights in the process and structure of action

with scientific insights in these, adding two other formats of representation of action. Our aim is to further add to this framework in order to develop it such that it can also account for the aristotelian observation that even moral action can over time be habituated or automatized and still not lose its moral significance.³⁰²

Therefore, we need a comprehensive framework that allows us to explain how an agent is capable of sculpting and employing a space of actions, partly determined by intentions and ideals to which he has committed himself earlier. Such a space of actions will allow him to act flexibly in a fast, stable yet also implicit manner, while still respecting some important constraints for action that he has taken upon himself. So we are interested in a framework that can explain how explicit intentions contribute to the formation of such a space of actions instead of maintaining a strict distinction between these explicit intentions and the space of actions. Only then are we allowed: “usher habitual actions, or at least a subset of them, into the space of reasons. That subset will consist of those habitual actions which cohere with the agent’s world view” (Pollard 2005 81).³⁰³ A couple of characteristics can be formulated that would hold for such a framework. The first two characteristics are related to the mechanisms and the representations involved, while the other two characteristics refer to structural properties of the comprehensive result of these.

A first characteristic refers to the rather complex nature of the mechanisms involved. Based upon our earlier observations, we are not only expecting to find the relevance of a hierarchical structure of control, but we also expect to see how a sculpted space of actions is being employed. This space of actions will likely contain action representations at several levels of specificity and will have been sculpted over an extended period of time. That is to say, even though there is an important role for the top-down flow of control, this does not imply that at all times an action must be determined through the actual involvement of the comprehensive cascade of intentions. For one, as we’ve noted several times earlier, the hierarchical structure of complex and dynamical systems is in fact heterarchical, allowing the development of direct connections of upper with bottom levels of the hierarchy, with the evasion of intermediate levels (Berntson and Cacioppo 2008). For another, with a sculpted space of actions at several levels of specificity in place, not at all times do actions

³⁰² Concurring with Aristotle, such habituation of moral action can be considered as the development of a ‘second nature’, leading to the automatization of its performance (McDowell 1994). However, what needs to be warded off in that case, is the critique that such an action does no longer have its origin in the ‘space of reasons’ but merely in the ‘space of causes’, making it rather comparable to a hard-wired reflex. McDowell argues that Sellars – who introduced the notion of the two spaces (Sellars 1997) – made a strict distinction between the two, which he aims to tear down (McDowell 1994).

³⁰³ With that, it can be argued, acting for reasons can even hold for those actions that we perform automatically or habitually, without necessarily always requiring preliminary reflection (Pollard 2003).

require still to be determined by the full intentional cascade. Perhaps only triggered by an unexpected orchestral intro the expert singer can switch from performing Don Giovanni to Saint François instantaneously, without needing to revisit his earlier reflections upon the interpretation of these roles while still using the corresponding modes of singing and acting. Instead, action and motor representations pertaining to these roles are already established and the appropriate ones can be activated through more or less consciously established intentions.

A second characteristic refers to the multiple representations that are involved in the determination of action. As we have learnt from the previous Part, in which we discussed development and learning, a single task can be carried out with the use of different representations, some of which are redescribed versions of others. In the present Part, this would imply that indeed more than just a singular representation of a particular action can be involved in the complex process leading to an action. Consequently, a challenge is the handling of these multiple representations, avoiding the influence of potentially incongruent representations. Or it may be difficult to specifically modify an action when it is performed habitually, for the underlying representation is then activated as a whole. It may be difficult for our expert singer to give his Don Giovanni a more androgyn pose, because he has over the years practised a rather masculine voice and pose, making it difficult to single out his pose for adjustment in another direction.

A third characteristic of the framework refers to the large productivity of actions that humans display and the differentiated role of certain action representations therein.³⁰⁴ Remember that we've been discussing earlier the fact that in complex and dynamical systems we can expect some components to become generatively entrenched (Wimsatt 1986), rendering these components a more prominent role in the system's performances than others. A similar observation was made in the previous Part concerning the kludges that are formed in cognitive mechanisms. As for the present context, we may expect to find that some actions are involved more generally than others in the configuration of novel action.³⁰⁵ Is it indeed the case that some actions or component actions, that are well practised and mastered, have indeed

³⁰⁴ Indeed, even though great apes demonstrate mastery of complex and hierarchically structured actions and thus a potential for productivity in their actions as well, it may be a lack of motivation and curiosity that keeps them from demonstrating such productivity (Byrne and Bates 2007).

³⁰⁵ Approaching this from a rather different perspective, Ricoeur has demanded attention for this capability in terms of the 'configuration' of action and the associated process of transfiguration that occurs both in performing and interpreting an action (Ricoeur 1991c ; Ricoeur 1992). Interestingly, Ricoeur also argues in favor of a certain hierarchy in the narration and consideration of action, which in turn can contribute to the organization and planning of life.

a greater probability of turning up in new configurations, of which the representations are figuring apparently more prominently in the agent's sculpted space of actions? Perhaps we can indeed find at several levels of specificity such deeply entrenched action representations, ranging from a particular vocal timbre to a more general mode of expressivity in an expert singer's performance, for example.

Finally, the framework's fourth characteristic refers to a structural property of the complex process of determination of an action. Given the diversity of component mechanisms and representations involved, a major concern could be how to secure a minimum amount of coherence and consistency between actions.³⁰⁶ Indeed, an important consequence of the hierarchy implied in the intentional cascade model is that it fosters the coherence and consistency between actions. The question that presents itself is whether an agent's sculpted space of actions does in fact enhance or endanger this property. For we may fear an expert singer can at times be relatively easily misled into an inappropriate performance if his attention is diverted to an irrelevant stage prop because of the unintentional activation of a part of his sculpted space of actions. Does this relative autonomy of an agent's sculpted space of actions thwart the intentional cascade's support his coherent and consistent acting, or is it perhaps difficult to decide about this?

Scrutinizing these characteristics, we will navigate between a philosophical account of intentional action and cognitive neuroscientific evidence about action selection processes, consisting of information about the representations involved and the neural implementations. While doing so, we will look for the leeway with regard to explicit specification of action that the philosophical account of intentional action to be discussed offers us, such that the empirical evidence regarding intentional action concurs with it.

³⁰⁶ This feature is related to the phenomenon of generative entrenchment, discussed in Part II. Once a stable feature becomes integrated in a complex mechanism, it is better able of developing – generating – new capabilities that build upon this entrenched feature, contributing to the mechanism's coherent performance (Wimsatt 1986 ; Wimsatt 2007).