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

Keestra, M.

Publication date
2014

3 DUAL-PROCESS THEORIES AND A COMPETITION BETWEEN FORMS OF PROCESSING

When invited to sing a particular character on stage, it may be difficult for a lay and an expert singer alike to suppress the many performative and vocal characteristics that we automatically associate with a juvenile and virile Don Giovanni or with a senior and meditative Saint François. Indeed, it is highly plausible that without further thought or consideration, their performance of these characters will conform to certain stereotypes and biases that are commonly attached to a young womanizer or to a senior monk, as can be seen in their movements and gestures and heard in their vocal expressions. Directions aiming to portray the character differently and in such a way as to surprise the audience will at first require careful attention of the singers, in order to inhibit their usual performance and adjust it accordingly. If the singer has already mastered the relevant score, it will give him more freedom to enforce these adjustments, as singing the notes does no longer require as much attention. Only then may his Don Giovanni express some fear instead of bravura when inviting the Marble Guest to dinner, or his Saint François express less serenity and more vitality than is usually the case. After playing a role several times under different directors, an expert singer will in a sense have different schemas or Gestalts of that character available, lingering somewhere in his memory and awaiting complete or partial activation. Unfortunately, when participating in a particular production again after some time, the singer may notice that the reactivation is not without partially confusing the current directions with other characterizations of the role. Nonetheless, it will usually take less time to reactivate the desired performance than it took him to initially learn it.

The behavioral and cognitive stereotypes that are evoked in the previous section are not just suspected to flourish in everyday life, but have been demonstrated in various experimental situations as well. Typically, such experiments use specific words or images as primes, which are thought to strongly activate particular associations. These prominent associations are usually stereotypes and biases that subsequently modulate the cognitive and behavioral responses of subjects, even if they would explicitly reject such responses. For example, being primed with sentences or images referring to elderly persons, subjects tend to use longer reaction times to respond to tasks. When primes refer to politicians, subjects tend to be more verbose, and when referring to fast animals, reaction times were shorter.[182]

[182] See (Dijksterhuis and Bargh 2001) for a review of these and other similar experimental results.

What these experiments have shown, is that our cognitive and behavioral responses often demonstrate the influence of implicit cognitive processes without the involvement of conscious deliberation. This does not come as a surprise, given our discussion of the phenomenon of modularization in the previous chapter, where we argued that modularization entailed an example of kludge formation. Complex behavioral and cognitive functions like playing the piano or reasoning about physics, accordingly undergo a double featured process. For one thing, the progressive modularization that affects these functions makes them more encapsulated and thus less influentiable by other functions. Furthermore, the explicitation process associated with the redescription of the representations involved points in an opposite direction, potentially making relevant information available for exchange, correction, articulation or other interactions (Karmiloff-Smith 1992). Based upon this insight, the strict distinctions between declarative and procedural, or between conscious and unconscious, or between controlled and automatic forms of processing have been contested, since the developmental relation that connects them apparently does not lead to the simple substitution of one form of processing by another (Karmiloff-Smith 1992 26). This nuanced position is not shared by all researchers of human cognitive functions, as we will see in this chapter. This has to do with the recognition that our cognition and behavior are not always driven by a single process only.

Indeed, since the argument of this dissertation relies partly upon the recognition that in many cases there is more than just a single process available for performing functions - even performing a complex function like intentional action - and that this facilitates the fast and flexible performance of such functions, we will further explore the validity of this position. In chapter II.3 we will focus on the 'dual-process theories', which appear to be in opposition to the nuanced insight mentioned in the previous section.[183] As mentioned, these dual process theories generally claim that the human mind functions according to (at least) two distinct types of processes, sometimes leading to conflicting outcomes. One type is a 'cognitive monster' that can hardly be controlled and automatically determines cognitive and behavioral performances in a prejudiced and stereotypical way, in contrast to the consciously controlled type of processing that is normally assumed to determine agents' behavior (Bargh 1999).[184] This latter mode, however, has much less influence on human performance than is

[183] This may well be due to the neglect of developmental issues in research concerning dual process theories, which has been acknowledged as an 'unfortunate omission' in a recent review (Evans 2011).

[184] Notably, even Bargh later acknowledged that forms of self-regulation or control of automatic processes are possible, suggesting that the 'dichotomy' between the two needs reconsideration (Hassin, Bargh et al. 2009).

generally thought, as our stereotypical opera singer exemplified.

We should in this context not lose sight of the fact that we should be applying different levels of analysis, again, to what perhaps appears to be a relatively simple problem. In chapter I.3, we discussed the three levels of analysis or explanation that David Marr distinguished: the task or computational level, the algorithmic level which also includes the representations used, and the neural implementation level (Marr 1982). Without additional research, just looking at a cognitive or behavioral response alone will not inform us what information processing underlies this response, nor will we be able to plausibly argue in favor of a particular neural network that carries out the required processing. Accepting these distinguished levels of analysis, our subsequent discussion in chapter II.5 of mechanistic explanation further explored how we can investigate a cognitive or behavioral task by decomposing it in component tasks which are tentatively further decomposed and located in a complex and dynamic mechanism. The dynamics of an explanatory mechanism was argued to permit certain modifications to occur, usually corresponding with modified properties of the cognitive and behavioral performance for which the modifiable mechanism is responsible.

Reminding ourselves of this background is useful, as we are embarking on a discussion of a set of dual-process theories that claim that a seemingly identical task can be carried out via two – or more, in some models – different processes, with some theories also presenting hypotheses concerning neural systems that are allegedly responsible for these processes. The methodological considerations that were brought back to mind in the previous section intend to emphasize that there is no straightforward relation between tasks, processes and systems. This lack of strict correspondence goes in either direction, as was argued in Part I: neural systems are often re-used or re-cycled for more than just a single process and task, while a particular task can be executed with distinct processes, probably relying on correspondingly different systems. Since dual-process theories focus on the hinging role that processes play in connecting tasks to systems, a further remark seems in order.

Dual-process theories are concerned with the fact that certain forms of information processing are more complex than others and correspondingly rely on different neural systems. However, not all tasks can be performed by more than just a single type of processing, which makes such a task more likely to be constrained by a specific implementation of the appropriate processes. This seems to be the case for visual information processing, where binding different features of perceived objects together is limited by the constraints of the underlying complex system – memory being part of that system (Treisman 1998). Nonetheless, expertise plays a role in perceptual processes as well, research on chess masters having shown that they are capable of recognizing

complex board positions within seconds, relying on many stored positions in their memory (Gobet and Simon 1996). On top of expertise in the sense of sheer amount of experience, another form of expertise will be seen to play a role in determining the form of information processing that an individual employs in a certain task: his ability to engage one of several representational formats of the task at hand. Indeed, a characteristic of many tasks besides visual information processing is that they allow different representations and thus potentially also allow different neural systems to be involved when they are carried out (Halford, Wilson et al. 1998).[185] As was mentioned in the previous chapter on modularization, changing a representational format of a certain task can greatly reduce computational demands while enhancing the possibility for learning, correction and generalization (Clark and Karmiloff-Smith 1993).[186]

In this chapter on dual-process theories, we will also discuss whether it is possible that the performance of a particular task shifts from one form of processing to another form. Although we already know from the previous chapter on modularization that tasks as diverse as playing the piano and mathematical reasoning allow such a shift, peculiar to the present discussion is that in many occasions two distinct forms of processing appear to compete for determining the task outcome. Since particularly one of the two forms is considered to be seriously impeded by its large computational demands, this competition is often won by the 'cognitive monster' mentioned above, as this monster proceeds differently (Bargh 1999). Let us look more closely at the account given of the two forms of processing involved and subsequently investigate whether a shift of processing is possible, entailing another form of kludge formation.

3.1 Distinguishing between forms of processing, irrespective of tasks?

The dual-process theoretical assumption that a particular task can be carried out via two very different types of processing was based upon research like the experiment in which subjects were required to perform the so-called Wason-task, after which the researchers asked their subjects to "write down your reasons for choosing to examine or to ignore" a particular feature of that task (Wason and Evans 1975 142). The authors

[185] Halford et al. assert that visual information processing is highly modularized, making it very hard to 'reprogram' it in a strategic way as can be done with higher cognitive processes (Halford, Wilson et al. 1998).
[186] Cognitive complexity is proposed as reflecting "the ability to comprehend a cognitive domain with a variety of independent attributes for describing the objects in it" (Scott 1962 410). Cognitive flexibility is then defined as the ability to change the representations of the objects within the domain by focusing on different attributes, issuing in different decompositions of the domain. More specific is the notion of complexity that refers to entities and relations within a domain, where a domain usually allows multiple descriptions and descriptions at different levels of abstraction. For example, "Relations in a familiar domain can be more readily chunked, or higher order relations may be known that allow the structure to be represented hierarchically" (Halford, Wilson et al. 1998 811).

conclude from the incongruence between the observed results and the introspective reports by the subjects that the performances and the introspective renderings of the task referred to different processes. Moreover, the biased reasoning that transpired in subjects' performance for the Wason-task appeared to be impenetrable or unavailable for their introspection. This rendered the introspective report delivered by the subjects post factum a mere construction or 'rationalization'. The researchers concluded from this observation that, contrary to common sense, at times a cognitive or behavioral task performance is not preceded by inferential reasoning, even though one would expect such reasoning to underlie the performance. On the contrary, such a performance is at times only followed by a rationalization which turns out to be erroneous and a mere construction, and should therefore not be taken to be reliable reports of subjects' actual cognitive processes (Wason and Evans 1975).

Since those early investigations, dual-process theories have been proposed as explanations for the conflicting cognitive and behavioral responses demonstrated by subjects in the context of many different functions, ranging from social cognitive functions like attitudes, affect, self-regulation, social influences and blaming the victim (Chaiken and Trope 1999) to cognitive functions like reasoning and judgment (Evans 2008) and to forms of motor behavior (Hofmann, Friese et al. 2009). Indeed, even domain-unspecific functions like memory and learning are being approached from this dual-process theoretical perspective (Frankish and Evans 2009). Common to all such examples of dual-process theories is their emphasis upon those knowledge representations that are usually learnt implicitly and unconsciously and subsequently determine cognition and behavior in a similarly implicit and unconscious way, even though agents tend to think that their cognition and behavior is largely driven by explicit and conscious information processing.[187] Indeed, notwithstanding their differences, most dual-process theories share several attributes.

Although not all relevant authors agree that the two distinct types of processes are served by equally distinct – cognitive and neural - systems,[188] it has become common to refer to systems 1 and 2 respectively even when two types of processing are in fact

[187] Dual process theories are not unlike the reinterpretive work of the Masters of Suspicion – Marx, Nietzsche, Freud – who also did not accept the explicit and conscious self-accounts of fellow authors but instead argued that their mistaken or alienated self-construals in fact hid other factors determining human culture and thought (Ricoeur 1970). Not using the expression, they would probably agree to calling these other factors 'cognitive monsters'.

[188] Terminology among dual-process theorists is somewhat confusing and also liable to change. Two key authors – Evans and Stanovich - appear to agree in now favoring reference to two different 'types of processing' instead of two systems, with these processing types interacting when producing mental performances as well (Evans 2011). As this terminology concurs with our argument that a plurality of processes is available for many tasks for agents, who can learn to get some control of these, we will adopt this reference to 'types of processing'.

intended.[189] Similarly, most authors agree on the sets of attributes assigned to these two types of processing, notwithstanding some remaining differences between authors. In a recent historical and systematic review, the attributes of the two systems – or types of processing - are listed in Table 1, adapted from (Frankish and Evans 2009):

The previously mentioned distinction between task performance, form of algorithmic or information processing, and neural processes or systems is somewhat reflected in this table. The reference made to evolutionary age and distribution of the systems is particularly valid for their neural implementation, irrespective of the particular form of information processing carried out by those systems. As mentioned earlier, the fact that human and animal brains share many structures and properties does not withstand the fact that such structures are exapted for different forms of processing in humans, as well (Anderson 2010). Not surprisingly, the table especially pays attention to the differences in information processing between the two types, in

System 1	**System 2**
Evolutionary old	Evolutionary recent
Unconscious, preconscious	Conscious
Shared with animals	Uniquely (distinctively) human
Implicit knowledge	Explicit knowledge
Automatic	Controlled
Fast	Slow
Parallel	Sequential
High capacity	Low capacity
Intuitive	Reflective
Contextualized	Abstract
Pragmatic	Logical
Associative	Rule-based
Independent of general intelligence	Linked to general intelligence

Table 1. Features attributed by various theorists to the two systems of cognition. Adapted from (Frankish and Evans 2009 15) with permission from the publisher.

[189] Once it is posited that two distinct neural systems underlie the two processes, a host of additional empirical hypotheses follow. For instance, dissociations between the two processes should be discernible in patients with lesions that affect one and not the other system. An early proposal for a two systems account assumed that the two processes recruit different memory systems, one being associative and the other rule-based (Smith and DeCoster 2000). More recently, neuroimaging results of experiments in which both processes are activated have led to the distinction between a reflexive, automatic system and a reflective, controlled system (Lieberman 2007).

other words to the 'algorithmic theory' in Marr's (Marr 1982) sense. The influence of rules or mere associations, the presence of abstract or contextualized information, the sequential or parallel nature of the informational process, and the difference in information load refer to these information processing differences. Associated with these processing differences are, finally, also observable differences in the task performance, such as its being conscious, its explicitability, its controllability and the influence of reflection on it.

Besides, and somewhat confusingly, Table 1 does not differentiate between attributes that refer to the learning process and others that refer to the activation of the previously learnt knowledge. This is most apparent with the pair 'fast – slow' which here refers not to the speed of learning but to the speed of activation, which happens fast for automatic and not for controlled processes. With regard to the speed of learning the pair would in fact have been the other way around, as conscious and rule-based learning can happen instantaneously, while unconscious and associative learning is dependent upon repeated exposure to the relevant information.[190] Finally, the table suggests that there are just two different systems or types of processes, while several authors argue that particularly system 1 or processing type 1 can be subdivided, with others arguing, conversely, that a single system underlies all different processing types depending upon the way it has been triggered by specific cues.[191] However, notwithstanding these differences there is agreement between most authors about two aspects, since: "[a]ll that really links dual process theories together is the nature of System 2 and the way in which implicit and automatic processes (of whatever kind) appear to compete with it for control of our behavior" (Evans 2006 205).

It is particularly the latter aspect that we will discuss in this chapter on dual process theories: the fight for control over cognition and behavior between the two

[190] This distinction between slow and fast learning has been aligned with distinctions between memory systems and types of content in the influential dual-process account presented in (Smith and DeCoster 2000). However, this content-based distinction between associative and rule-based learning has been challenged, as associations can also be taken to be a particular type of 'if-then' rules (Kruglanski and Orehek 2007). This has been defended even for the case of conditioning (Holyoak, Koh et al. 1989).

[191] Stanovich, for example, refers not to a single system 1 process but to 'TASS': The Autonomous Set of Systems. Common to these TASS is that they are fast, automatic and mandatory. However, he emphasizes that some TASS processes or particular goal states of these TASS processes may in some cases become automatic only after practice, which is not commonly attributed to System 1 processes (Stanovich 2005). Elsewhere, he distinguished the non-TASS processes in a 'reflective mind' and an 'algorithmic mind' the first signaling the need to employ non-TASS processes to a certain situation, the second then carrying out a reasoning task (Stanovich 2009). A different model distinguishes two automatic and two controlled processes. In this Quadruple model, the four processes may interact in a single task, depending on context, response tendencies, information availability and other task features. Consequently, the model allows the subject to engage in various forms of self-regulation and self-control (Sherman, Gawronski et al. 2008). In contrast, Kruglanski et al. propose a uni-model that responds differently to specific parameters of the task at hand (Kruglanski and Orehek 2007).

types of processing – which we will refer to from now on as automatic and controlled processing, respectively.[192] This fight for control is intimately linked to the distinctions in information processing together with the differences in the representations involved. One way of performing a task can involve a more comprehensive representation of the necessary information than another way of performing it. Associated with differences in representation format are, obviously, differences in information processing that rely on neural systems that can differ in kind and in number. The task of squaring the number 6, for example, is for many subjects a matter of activating their memorized table of 6, while others may have to add 6 sixes – which involves reliance on the capability to add and on working memory.

As we are more specifically interested in the process of kludge formation as it may contribute to a more stable 'sculpted space of actions', we will inquire to what extent is there a shift possible between the two: can a type 2 – controlled - process itself become automatic and thus gain more control over someone's cognition and behavior? Or is the line between the two types of processes strict and non-permeable, leaving automatic processes largely immune to the interference by a controlled process?

3.1.1 Considerations of the distinction between automatic and controlled processes

Before considering the possibility of automatization of type 2 processing, let us ward off the objection that type 2 processing can *per definitionem* not become automatized, as such a shift would render the conceptual distinction between automatic and controlled processes meaningless. There are at least three possible responses to such an objection. First, by making a gradual instead of a strict distinction between conscious and unconscious, or between implicit and explicit, or between automatic and controlled processing, we are better able to account for empirical and computational results. Notwithstanding the gradual nature of these distinctions, we can still recognize different phases with their own specific properties, for example with regard to the representations involved (Cleeremans and Jiménez 2002).[193] Second, as we noted in the previous chapter on modularization and the corresponding processes of proceduralization and explicitation, observable behavioral effects of

[192] Obviously, as is the case with a concept like 'implicit', the concept 'automatic' can be further decomposed in several features like unintentional, uncontrolled/uncontrollable, goal independent, autonomous, purely stimulus driven, unconscious, efficient, and fast. After analyzing these features, it is argued that the distinction with non-automaticity is gradual, rather than strict (Moors and De Houwer 2006), which concurs with our argument further below.

[193] Indeed, the previous section yielded the insight that learning – which was associated with a gradual yet multistage modularization process - can even result in: "the existence in the mind of multiple representations of similar knowledge at different levels of detail and explicitness" (Karmiloff-Smith 1992 22).

learning particularly concern the shift along precisely these gradients of a particular task (Karmiloff-Smith 1992). These effects, therefore, affirm that it makes sense to distinguish between these forms of processing, even though they are connected in a developmental or learning trajectory.[194] Third, specific to most dual-process theories is the assumption that a subject's actual cognition or behavior is the outcome of a competition between two different types of cognitive processing which differ particularly in the conditions of their activation.[195] Automaticity in this context does not so much refer to a strictly distinct type of processing but primarily to the activation of processes due to external triggers not necessarily selected by the subject. Controlled processing accordingly refers to the internal or intentional activation or – so we will argue – to the internal or intentional selection of those triggers that eventually activate cognitive processing.[196] Having considered these arguments against a strict conceptual separation of automatic from controlled processes, let us then proceed with the main question: what benefits should we expect to stem from a shift from controlled to automatic processing of a particular task? To answer this question, it is important to note the limitations that affect the causal or determinative power of controlled processes and to consider how these limitations are related to the kind of task-dependent information involved in these processes. Those limitations primarily concern the neural underpinnings of the processes that carry out the tasks and are therefore only in a derivative sense related to the information that is processed. As dual-process theories generally share the conviction that the process limitations

---

[194] In this context it may be noted that Aristotle's introduction of the dynamis-energeia gradient has provoked as much debate as it has helped thinking, particularly in the life sciences and medicine. The ancient resistance and the modern difficulty with similar notions may be related to a philosophical and – meanwhile – scientific preference for mathematics and physics with their more strict distinctions or fixed relations between entities, be they magnitudes or particles.

[195] The competition between processes can be configured differently. For example, two processes may simultaneously seek to control the outcome of processing, or one may seek to overcome or correct the preceding process, or a third process may select one of two processes to determine the outcome (Gilbert 1999). An account that has proven to be fruitful not only in explaining experimental but also in predicting results of training is a relatively simple model in which impulsive and reflective systems compete for the final determination of cognition or behavior (Hofmann, Friese et al. 2008 ; Strack and Deutsch 2004). It should be noted, however, that not all models assume such a competition and that other configurations are possible between the systems (Gilbert 1999). Interaction of systems, assumed to simultaneously contribute to a task performance, is assumed in accounts of (Cunningham, Zelazo et al. 2007 ; Smith and DeCoster 2000 ; Stenning and Lambalgen 2008 ; Sun, Slusarz et al. 2005)

[196] In contrast to the suggestion that controlled processing by definition cannot be automatized, one may thus doubt the value of a strict distinction between the two (Bargh and Ferguson 2000). Concurring with the latter is the observation that controlled, goal-directed attention allocation may influence subsequent automatic processing, which interplay of forms of attention: "may obfuscate the need for the distinction between automatic and controlled processing whatsoever" (Feldman Barrett, Tugade et al. 2004 567). Nonetheless, as is generally the case with terms that refer to functions or processes that are dynamically related, it may still be useful to distinguish them when there are empirically distinguishable properties involved.

matter most and that the limitations of controlled processing are responsible for the prominence of automatic processing, we will first approach that issue.

3.1.2 Processing limitations held responsible for the distinction between automatic and controlled processing

Note that the table discussed in the previous section lists as properties of controlled processes among others their being slow, their low capacity and their dependency upon consciousness (Frankish and Evans 2009). Indeed, it is such limitations in capacity of controlled processing, in view of the ongoing and multiple demands of a subject for immediate responses to his volatile environment, that would allegedly make some shift of the information processing load to automatic processing desirable. As automatic processing is fast, has greater capacity, can occur in parallel and does not need to involve consciousness, it is considered to be the default type of processing, leaving to controlled processing a limited role which can only be deployed sparingly.[197] Obviously, once controlled processing of a certain task could also become automatized, task performance would perhaps no longer be subject to the limitations that hold for controlled processes. In our discussion of the dual-process theories we will argue – particularly in section II.3.1.5 - that the complexity of a particular task should not be considered as a static fact. Instead, there are strategies available for reducing the complexity of the information that requires processing for a particular task, comparable to the process of Representational Redescription discussed in section II.2.1, that is involved in children's mastery of certain skills. Consequently, not only is complexity to some extent adaptable, it is also problematic to determine capacity limitations in terms of information processing limitations. Nonetheless, we will consider some dual-process accounts that focus on a particular processing factor or neural component as being responsible for such limitations.

Authors do not completely agree in their diagnosis of the most important bottleneck that yields this limited capacity of controlled processing. This disagreement may be partly due to differences in focus on one or another component of a complex and dynamic mechanism for information processing and decision making, where this mechanism may not at all times engage all of its components with equal strength. However, as both processing types are involved in information processing, the capacity limitation at stake makes itself felt particularly with regard to the *quantity*

[197] Considering different prominent obstacles for a dominance of controlled processing – Reflective system processing, in their terms – like time pressure, cognitive overload and alcohol intoxication, Hofmann et al. conclude that all of these obstacles are related to impediments of working memory (Hofmann, Friese et al. 2009).

of information that is being processed. Where is this informational bottleneck to be located in the mechanism – and is there a single bottleneck, or are there potentially several components involved?

Some authors identify consciousness as the bottleneck. As was visible in Table 1 above, most dual process theories consider the type 2 processes to be conscious. The 'unconscious-thought theory' appears to locate the shortcomings of conscious thought in dealing with complex problems in its limited capacity for dealing with large amounts of information (Dijksterhuis and Nordgren 2006).[198] Nonetheless, consciousness is here not defined in such a way that it clarifies why consciousness should present the bottleneck nor how we should determine its capacity.[199] Instead, the unconscious-thought theory refers to Miller's account of the informational limit that has been so influential since his article on "The magical number seven, plus or minus two" (Miller 1956). Interestingly, that account focused on the capacity of memory, and not on the capacity of consciousness.[200] Indeed, it is questionable whether consciousness can be held responsible for the capacity limitations.[201] Perhaps, therefore, this alleged limited capacity of consciousness could be a downstream effect of memory – both being part of a more comprehensive information processing mechanism. Let us first consider whether we can explain differences in cognitive and behavioral responses, following the suggestion that different systems are involved in types of information processing, like different memory systems.

---

[198] Unconscious-thought theory is usually distinguished from dual-process theories in that it does not assume the presence of two different systems – though this is not a strict prerequisite for dual-process theories, as we observed above. Furthermore, the theory differs from common dual-process accounts in that it does not criticize but highlights the value and optimal outcomes of certain effortless, unconscious thought processes. Crucial, however, is its denial of a strict distinction between conscious and unconscious thought with respect to the kind of input – associative or rule-based – that is used (Dijksterhuis and Nordgren 2006). Our account of a shift from controlled to automatic processing concurs with that denial.

[199] In another article, conscious thought is defined as 'deliberation-with-attention', referring now to attention as a limiting component (Dijksterhuis, Bos et al. 2006). The authors there suggest that doing arithmetic requires conscious attention, failing to acknowledge that doing arithmetic consciously and attentively yields extremely divergent results in novices and masters, even when applying identical arithmetic rules.

[200] Miller underscores the importance of recoding – particularly linguistic recoding – of information that helps humans to counter to some extent the limitations of memory (Miller 1956). This is not denied in the unconscious-thought theory, even though consciousness is being denied a functional role with regard to determining the productive information processes via encoding and recoding of information (Dijksterhuis 2007).

[201] Such an association of consciousness with computational capacity limitations is explicitly denied in a version of the 'workspace' theory of consciousness, which equally rejects the notion that consciousness corresponds with particular neural systems. Instead, the authors argue that it is the specific type of neural activation that is responsible for making a particular content consciously available (Dehaene and Naccache 2001). The workspace theory of consciousness even allows specialized computational processes to run unconsciously in parallel, each being responsible for making only relevant parts of the overall task consciously available, while conscious processing occurs serially at the same time (Shanahan and Baars 2005).

3.1.3 Memory systems invoked for the explanation of the distinction between automatic and controlled processing

Indeed, a distinction can be made between two different memory systems, which are also involved in different forms of learning: "a memory system that supports gradual or incremental learning and is involved in the acquisition of habits and skills and a system that supports rapid one-trial learning and is necessary for forming memories that represent specific situations and episodes" (Sherry and Schacter 1987 446). It is the latter memory system, playing a role in many cognitive functions as working memory - its central executive component[202] in particular – and its limitations that essentially impedes controlled processing. Many of the differences that have been found in individuals with regard to the interaction between the two types of processing may be attributed to differences regarding their working memory.[203] As a result, an individual can be called either a 'motivated tactician' with high working memory and consequently large controlled processing capacity, or a 'cognitive miser' with low capacity and therefore with a relatively larger role for automatic processing (Feldman Barrett, Tugade et al. 2004). The question then presents itself whether subjects can only demonstrate cognitive and behavioral responses according to either one of these types, or whether they can to some extent shift between these two types.

Several authors focus on the role of memory for answering this question on shifting between processing types as a way to modify responses. The assumption is that subjects need to sculpt their space of actions by bringing about such a shift in processing. According to the so-called Reflective-Impulsive model, Type 1 or automatic processing is carried out by the impulsive system which 'slowly and gradually' yields a network of associative connections related to behavioral schemas. The result can be considered to be a 'conceptual and procedural long-term memory' that can process large amounts of information in parallel, with activation of multiple response options as a result. In contrast, Type 2 or controlled processing activates behavioral schemas by way of the reflective system. Controlled processing relies on executive control and working memory and is accordingly impeded by severe capacity limitations (Deutsch and Strack

---

[202] In the words of a pioneer of working memory research, the central executive is still: "the most important but least understood component of working memory" (Baddeley 2003 835). Nonetheless, in his review Baddeley associates working memory with forms of self-control and regulation and contrasts this with implicit forms of control which do not rely on working memory – concurring with the dual process approach discussed here.

[203] However, it would be mistaken to assume that automatic processing which issues in a response does not engage working memory, just like it is a mistake to assume that controlled processing in no way involves automatic determination of components of the response – like in implementation intentions (see more on the latter below) (Bargh and Ferguson 2000). So even though working memory limitations may make themselves more notably felt in controlled processing, working memory is involved in automatic processing as well.

2006). According to the model, both types of processing activate in their separate ways behavioral schemas, and a competition between these schemas eventually leads to the execution of a particular behavioral schema.[204]

However, even though the distinction between controlled and automatic processing is perhaps associated with differences in the recruited memory systems, it cannot be associated with a strict distinction between tasks. That is to say, one and the same task can be performed according to different strategies, relying on different task representations and as a result, can also be associated with different – neural – systems. For example, applying the Reflective-Impulsive model, research has demonstrated that with repeated conscious and intentional exercise of a particular behavioral schema, this schema will over time be processed with the impulsive instead of the reflective system. As a result, preferred and self-regulated behavior can become part of automatic processing, thus avoiding the limitations of controlled processing (Hofmann, Friese et al. 2008). Indeed, in contrast to the sometimes worrisome reflections on the prevalence of biased and stereotypical responses that are raised in the context of dual-process theories, room is left open for a process of 'sculpting the space of actions' even in this theoretical context.

3.1.4 Some strategies that allow a shift between automatic and controlled processing

The repeated practical exercise of self-regulated behavior as a way of modifying the mechanism responsible for our behavior or cognition generally implies a specific role for controlled processing. However, this is not the only strategy available to agents when they seek to change their behavioral or cognitive response patterns, upon the recognition of their limitations of controlled processing. Other strategies are available, differing in nature and in the focus on the action component needing adjustment. This section will mention a couple of such strategies, merely to show the variety and to demonstrate that kludge formation can affect responsible mechanisms.

To begin with, automatized self-regulation need not just pertain to behavioral properties, like the goal of an action. It can also be directed at emotions associated with actions. Several strategies have been shown to be effective in such automatic self-regulation. For example, an agent can prepare himself by formulating so-called implementation intentions that concern a particular future action in a specific

[204] A different approach to the distinction between subjects' accustomed behavior and their goal-directed actions uses the construct of habit as environmentally cued behavior. This construct, too, recognizes the interaction between the two processes which enables agents to try to mitigate undesired habits (Wood and Neal 2007).

situation, which are shown to influence his future actions even if these are done automatically (Gollwitzer and Brandstatter 1997). In a similar vein can an agent prepare himself by articulating emotions that he considers to be appropriate or he can practice to automatically withdraw his attention from negative stimuli or to avoid negative interpretation biases (Gyurak, Gross et al. 2011).

More common are strategies that are directed at the outcome or goal of an action. Apart from practice-based modification, an agent can also activate consciously and intentionally a preferred behavioral schema with controlled processing and thus prepare for future situations by alleviating the taxation of controlled processing during those situations. Such a strategy makes use of agents' capability of automatic goal-pursuit, relying on the memorized representation of a habituated goal-directed action, which can be activated not only consciously but also without awareness (Bargh, Gollwitzer et al. 2001).[205]

For such unconscious goal-directed responses to occur, an agent can engage in the explicit formulation of intentions for specific goal-directed behavior under particular conditions. With such implementation intentions, an agent can obtain results that are comparable to habituated responses, even though the two strategies differ (Aarts and Dijksterhuis 2000). With the articulation of explicit 'if-then' rules for action, agents can anticipate future situations such that they act more reliably in such situations according to their previously formulated intentions. Yet they respond to a present situation with the intended behavioral response in an immediate and efficient way, without the involvement of conscious intent (Gollwitzer and Sheeran 2006).[206] Indeed, what has been formulated via a controlled process has become automatized such that the intended behavioral response will be automatically activated by the anticipated situational cues (Webb and Sheeran 2007).[207]

While the implementation intention strategy relies on the preliminary activation of a behavioral schema, thus facilitating its execution at a later stage, an alternative for agents is to engage in counterfactual thought as a self-regulation strategy in order to

---

[205] In spite of earlier convictions that automatic goal pursuit is by nature inflexible and irresponsive to altered environmental or internal states, more recent research has demonstrated that automatic goal pursuit can under circumstances in fact be flexible and responsive (Hassin, Bargh et al. 2009). Agent's preparation can play a role in this, as can the representation of the task, as we will argue below.

[206] These three features, immediacy, efficiency and lack of conscious intent, are considered prime attributes of automatic processes, as the authors note correctly (Gollwitzer and Sheeran 2006)

[207] Indeed, one of Kruglanski's reasons for developing a uni-model in which the dual processes are unified is that the associations that determine automatic processing can also be considered as if-then rules – even though they are not explicitly and consciously formulated by an agent during his response (Kruglanski and Orehek 2007). Lieberman makes a similar admission and has an equal dislike of the strict separation of automatic from controlled processing, inspiring him to the development of yet another model in which interactions between these processes are prevalent (Lieberman, Gaunt et al. 2002).

diminish the chance of executing an undesired action. Such simulation or imagination of future, hypothetical response options is typically associated with controlled processing (Evans 2008). In this case, remarkably, activation of a response option through controlled processing influences subsequent automatic processing under specific conditions. For example, counterfactual thought, devoted to an alternative course of action or judgment while considering a past event, can be used to prime a future cognitive or behavioral response and leads to less biased responses (Galinsky and Moskowitz 2000). Consisting of the activation of memorized representations of previous events and then on the simulation of alternatives – or 'variations on a theme' -, counterfactual thought depends on cognitive and neural processes that are similar to those involved in normal action planning and coordination. (Narayanan 2009).[208]

Now that we have considered some different strategies for automatic self-regulation, a few remarks should be made. First, these strategies and the shift from controlled to automatic processing that they bring about, once again emphasize that it is implausible to strictly separate the two types of processing. Indeed, one of the reason for proposing the Quad-model as an alternative, is to account for the various forms and focuses of self-regulatory strategies that agents have at their disposal (Sherman, Gawronski et al. 2008). Second, many strategies in some way involve a representation of a relevant action or action situation.[209] Such a representation can refer to an action at different levels of grain or abstraction, with different regulatory outcomes. Given the relevance of this aspect of representation, which will be more at the focus of Part III, let us pause for a moment to consider such task representations, or the 'algorithmic' theories (Marr 1982) involved.

3.1.5 Representational differences and the shift between automatic and controlled processing

To summarize the previous section, both the intentional adaptation of a practice-based

---

[208] Counterfactual thought allegedly exploits the representation of 'Structured Event Complexes', which generally have a hierarchical structure and a temporal sequence. Based upon the analysis of these SEC's, counterfactual thought is assumed to be most effective when focusing on the central dimensions of action-inaction, self-other, and event outcomes (Barbey, Krueger et al. 2009).

[209] There is a heated debate about the notion of representation in cognitive neuroscience. The debate concerns questions such as whether we need to involve representations at all, or should instead employ notions from dynamical systems theory, for example (Keijzer 2002). A related topic is whether the notion of representation adequately captures the properties of distributed information in neural networks, influencing network activations. (Bechtel 1998) defends this position against among others (van Gelder 1998). Given the fact that self-regulation can be successful with the preliminary invocation of an action representation, we consider the notion both plausible and useful here. This concurs with the observation that representations are involved in forms of learning, transfer and correction of cognition and behavior (Clark and Karmiloff-Smith 1993), as is also the case in so-called 'representation-hungry' tasks involving distal, non-existent or highly abstract action properties (Clark and Toribio 1994).

strategy and the strategies relying on the explicit formulation of specified intentions or of counterfactual intentions, therefore, are capable of modifying the agent's mechanism that is responsible for his behavioral response in particular situations. Obviously, the modifications obtained with these strategies are not equally structural or long-lasting. Indeed, further down – in chapter II.4 - we will discuss further strategies or tools available for such modifications that involve instruments or representations, linguistic and otherwise, that are even external to the skull in which our information processing takes place. Nonetheless, such modifications of 'internally' represented information do result in some change of an agent's sculpted space of actions, corresponding with some shift from controlled to automatic processing. This shift can occur for actions that conform to a more or less specified representation of a cognitive or behavioral response. As a result, the limitations in capacity of automatic processing are circumvented, notably the limitation of working memory that is crucial for controlled processing. Indeed, having presented his influential account of the limitations of memory, Miller emphasized that a recoding of information can help to mitigate these limitations (Miller 1956). Fortunately, therefore, the distinction between the two processing types is neither strict, nor static, allowing a specific controlled process to shift to automatic processing, thus alleviating the burden on working memory.

In the next section we will close our discussion of dual-process theories with a consideration of the seven kludge characteristics, which we use as a means to estimate the agreement of these theories with the observation that modification of an explanatory mechanism responsible for an agent's action can occur. Before engaging with this evaluative task, we'll devote this section to a discussion of the task or information processing involved in both automatic and controlled cognitive processes. We already know from our earlier discussion of the modularization process in development and learning that the representation involved in certain cognitive or behavioral tasks does matter in an important way. Both the proceduralization and the explicitation that were found to be associated with the phase-wise mastery of a cognitive arithmetic task or with a behavioral task like playing the piano were argued to rely on a process of 'Representational Redescription' (Karmiloff-Smith 1992). More generally, the human capability of employing different formats of representation for the performance of a certain task enables humans to engage in corrections, generalizations, and transfer more than animals do (Clark and Karmiloff-Smith 1993).

How can we extend this insight to the present issue of two allegedly distinct, yet competing processes that are potentially involved in performing the same task? As we noted regarding Table 1 on the two cognitive processes in section II.3.1, some of the attributes included there refer to the differences in the representations involved in

those processes. However, these attributes should not seduce us into thinking that a strict distinction is at stake, although the table associates the sequential processing of rules and abstract information with controlled processing while leaving for automatic processing the parallel processing of associations of contextualized information (Frankish and Evans 2009). Instead, we will present a couple of arguments why such a strict distinction between the tasks or representations involved must be rejected.

To begin with, in contrast to the distinction between associations and rules mentioned in the table, associations can be considered to be a particular type of 'if-then' rule (Kruglanski and Orehek 2007). Indeed, even classical conditioning of behavior, typically depending on association learning, has been defended as being a matter of learning to follow such a rule (Holyoak, Koh et al. 1989).[210] Comparable with this identification of association learning with a kind of rule-learning is also the 'Representational Redescription' process referred to above, when more abstract re-representations can arise spontaneously after repeatedly performing a task.[211] Such a process can occur as a consequence of performing embodied actions, resulting in rule-learning in children and yielding observable results without involvement of verbal, conscious and controlled processing (Boncoddo, Dixon et al. 2010). Again, as argued earlier, controlled and automatic processing should not be placed strictly apart, nor should we assume the information involved in these types of processing to be strictly different. Let us pursue this last point somewhat further.

Defining the complexity or computational load involved in information processing has turned out to be an intricate problem, which we will not try to solve here.[212] Our more modest aim here is to argue that – irrespective of its specific definition - the complexity of a task is not a static given but allows some adjustment. For example, a particular approach to this issue starts from the analysis of complexity as *relational complexity*: not just the number of items implied in a certain task matters, but more specifically the number of arguments or relations between them (Halford, Wilson et al. 1998).[213] Complexity then increases with the number of interacting variables, but

---

[210] It is usually held that relatively lower level cognitive processes, like visual perception, are modifiable merely through learning simple association rules. In contrast to this, research demonstrates that expertise with rather abstract mathematical reasoning does also influence early perceptual processing, as experts visually group notational elements in mathematical exercises different from novices (Goldstone, Landy et al. 2010). We will return to this in section II.4.2.1 below.

[211] Such learning could be described as self-organization of the relevant neural networks, enabling the emergence of new properties – in this case, new representations (Stephen, Dixon et al. 2009).

[212] There are several different definitions of complexity available, for example with regard to the question whether or not evolution leads to increasing complexity (Chu 2008). Preliminarily, however, one would need to decide what the valid comparisons are in this context, what time scales are to be applied, and how the complexity is operationalized and measured – which is difficult without a definition (McShea 1991).

[213] A different yet comparable account focuses on the number and the complexity of the relations that are used in task performances and is equally interested in subjects' flexibility in using a different set of relations to reduce a task's complexity (Zelazo and Frye 1998)

especially when the number of different relations between these variables increases. When a series of even numbers needs to be added together, the complexity is less than when addition and multiplication must be alternated, for example.

However, such relational complexity can be transformed. Consequently, however it is quantified, complexity is not static and subjects can employ different strategies to reduce the complexity of information involved in a particular task. Such strategy change generally modifies a subject's performance and can yield different benefits, as each strategy activates different representations.[214]

One strategy is to segment a complex task in smaller component tasks, which can then be processed serially, another is to reduce complexity by chunking or collapsing the information into other relations. In that case, a hierarchical structure of relations is developed. The latter strategy is dependent, obviously, on two types of familiarity: "Relations in a familiar domain can be more readily chunked, or higher order relations may be known that allow the structure to be represented hierarchically" (Halford, Wilson et al. 1998 811). Our argument is not that with such strategies all processing limitations can be overcome, but merely that the task complexity can be reduced and consequently processing can be facilitated. Besides, we should remind ourselves what we learned above, that a redescription of representation can lead to loss of some information, as it often involves an abstraction of information. Or it might involve an exchange of information: for example discarding some information that is apparently irrelevant for the task at hand in favor of a different yet simpler representation.[215]

Indeed, other strands of research have confirmed that cognitive processes generally represent information at different levels of abstraction. In such cases, chunking may be a useful strategy. For example, the perception of unfamiliar actions by children already involves such hierarchical encoding, as they encode a complex action at several levels of abstraction simultaneously – encoding both a complex action and the component actions which make it up (Baldwin, Andersson et al. 2008). Indeed, subjects automatically segment perceived environments and actions according to a hierarchical

---

[214] According to this approach: "working memory is the workspace where relational representations are constructed and it is influence by knowledge stored in semantic memory" (Halford, Wilson et al. 2010 499).
[215] One critique of Halford's et al. target article argues that 'skill theory' provides a different account of learning to process higher relational complexity by: "(a) learning two simpler skills or networks, (b) mastery to allow parallel sustaining of the two, and (c) coordinating them in a new relation through multiple steps specified by skill/network combination rules" (Coch and Fischer 1998 835). The nature of the new relation requires some scrutiny. Not dissimilar is the critique from Cognitive Complexity and Control theory, which also emphasizes that some relation types may be more difficult to master than others, predicting differences in developmental trajectories dependent upon the relation types involved (Frye and Zelazo 1998). Notwithstanding their criticism, these authors still concur with the importance and prevalence of different complexity reduction strategies.

structure. In such cases, prior conceptualization of an action to be observed does not always determine the subsequent encoding of the observed action (Zacks, Kumar et al. 2009). Nonetheless, dependent upon a task to be performed, for example imitation, subjects are capable of adjusting the level of encoding of these perceptions to the task at hand. Verbal description can facilitate such encoding at several levels of hierarchy, but such differentiated encoding occurs as well after repeated observation (Hard, Lozano et al. 2006). The fact that such hierarchical encoding is not dependent upon conscious and language-dependent processing is supported by the fact that with their imitation learning capabilities, primates can be seen to engage in it as well (Byrne and Russon 1998). Here again, familiarity with a domain and with pertinent rules or hierarchical representations facilitate cognitive processing.

Before summarizing the present discussion, we need to engage with a related issue which can be illustrated with our example of expert singing. We emphasized that an expert singer is capable of modifying his performance in accordance with specific directions or intentions. Apparently, so we argued, an expert not only has flexible control over his performance, he can also access his performance via verbal or other cognitive strategies. This seems to contradict an alternative account of expertise (Dreyfus and Dreyfus 1986), which has been developed further since. According to this account, skillful performance does not rely on representations at all, implying that a skilled performer responds to environmental information without the invocation of explicit representations: "Unlike deliberate action, skillful coping turns out to have a world-to-mind direction of causation" (Dreyfus 2002 380) which allegedly leaves no room for representations of the world that are stored somehow in the mind.[216] Indeed, where a person engaged in improving his performance may rely on analytic reasoning, once skilled, this person is held to rely on intuitive decision making, which apparently rules out the contribution of representations (Dreyfus 2004). The implicit and automatic way in which an expert responds to his environment while skillfully coping with it, is taken to imply that representations no longer play a role in his cognition or behavior.[217] The shift from explicit and controlled performance of a task

[216] Dreyfus' phenomenological critique of cognitive science is connected to his phenomenologically inspired critique of the important role it assigns to representations. In that, however, he differs from Wheeler, who equally aims to develop a phenomenological account of cognition and behavior that sits well with cognitive science, yet leaves a role for representations. Importantly, these representations need to bear relevance for the agent, according to his account (Wheeler 2010)

[217] An ethical implication of this account of skill acquisition developed by Hubert and Stuart Dreyfus is that experts cannot be required to articulate and explicate their intuitive decisions, as they don't make these along the analytical, reasoning processes as novices do (Selinger and Crease 2002). This alone seems a very problematic aspect of the account, even though it may not be an inevitable implication. Holding experts morally responsible for their actions requires them to be capable of offering representations of these. Given our inclination to demand this of experts, it seems that our moral intuition does not concur with the account of intuition that the Dreyfuses provide.

to an implicit and automatic – in sum: intuitive – performance would correspond with a more holistic and non-analytic approach to a situation which must be explained in terms of neural dynamical systems and not in terms of representations, this account holds (Dreyfus 2009).

Although we risk reducing this discussion only to the potential role of representations, there are a few aspects worth mentioning in response to this account. First, in a critical response to this intuition-based account of expert skill, it has been pointed out that there is a lack of empirical evidence for the five distinguished stages and for the absence of analytic reasoning in expert performance (Gobet and Chassy 2009). Second, once we give an account of expert skill that involves an important role for the chunking of information, it implies that an expert is capable of capturing and processing strongly associated pieces of information at once. This is different from simply listing details as novices may need to do, simultaneously offering a way to account for an expert's difficulty of articulating his expertise (Gobet, Lane et al. 2001). This would require the articulation of chunked information that is still a representation, though it has meanwhile become a redescribed or re-represented one. Both neuroscientific and simulation studies suggest that this chunking process is even more complex than mentioned above. Indeed, chunks appear to have some hierarchical structure, as they function as templates with slots being left open for variable environmental information. As such, it can explain not only that experts are very quick in responding to complex environmental information – like in playing chess or in nursing – but that they do so while simultaneously responding flexibly to changes in detail of the environment (Gobet and Chassy 2009). Third, such chunks can become associated with emotional features, explaining why experts often respond with strong affective tendencies to a certain situation (Chassy and Gobet 2011).[218] Finally, without an account of expertise that involves the representation of information, it will be problematic to account for experts' capability of quickly learning, correcting, modifying, transferring and generalizing specific performances of their skill (Clark and Karmiloff-Smith 1993). Indeed, in our example of expert singers who can modify their performance of a Don Giovanni or other character when receiving explicit instructions, we can witness how representations play a crucial and effective role even when an expert usually relies on implicit and automatic processes for his performance.

After this short excursion, let us summarize the previous arguments. We have argued that, not just on the basis of considerations regarding automatic and controlled

[218] Building upon Dreyfus' account of expertise and adding to it the notion of 'action readiness' that stems from Frijda's theory of emotions (Frijda 1986), Rietveld has argued that an agent's concern-full, unreflective action is partly determined by the affective response that environmental affordances evoke (Rietveld 2008).

processes or their underlying systems, but also based upon the consideration of the information or representations involved, a strict distinction between controlled and automatic processes is not warranted. Instead, interactions between the processes should be expected. Indeed, a neural network model containing different levels of both automatic and controlled processing and assuming an iterative interaction between the two types of processing, turns out to be plausible both regarding system requirements and in predicting behavioral results (Cunningham, Zelazo et al. 2007).[219] Therefore, apart from the fact that shifts of processing do occur in various ways, there are also other reasons for not keeping the two types of processing strictly apart.

It is now time to wrap up the observations regarding the two types of processing by considering whether there is kludge formation at stake in this context, too. For we are interested in this part in modifications of an agent's mechanism such that we can observe some changes associated with his performance of a task, while other aspects remain the same.

3.2 Automatization of controlled self-regulation and the seven kludge characteristics

A kludge is assumed to be a part of an explanatory mechanism, being involved in the modification of such a mechanism. Indeed, that modification is to a large extent due to the formation of a new kludge, which obviously has some consequences for the mechanism, its properties and its observable behavior. In this section, we consider whether we can apply the notion of a kludge to cases in which a shift occurs from controlled to automatic processing. When development or learning has led to the establishment of a kludge, its first characteristic – presented in section II.1.1 above - reads that it should be recognizable on the basis of functional rather than other properties. If we refer once more to Table 1 included in section II.3.1, listing attributes pertaining to controlled and automatic processing, it clearly present mainly functional differences between the two, like speed, capacity, efficiency, explicitness, involvement of consciousness, contents, and so on (Frankish and Evans 2009). Indeed, comparable models, distinguishing between reflective and impulsive (Strack and Deutsch 2004) or between reflective and reflexive systems (Lieberman 2007) agree in that they distinguish between types of processing and make similar functional distinctions between the two. In the case of a shift between types of processing as a result of automatized self-regulation, for example, we can observe this primarily on the basis of the response

[219] A similar – bidirectional - interaction between the two types of processing can also account for social responses, depending upon the excitation or inhibition of specific representations (Adolphs 2009). See note 195 above for additional comments on different configurations between the types of processing or systems.

properties. Similarly, an incomplete shift or the involvement of both processing types is determined equally on the basis of observable cognitive and behavioral responses (Rydell, McConnell et al. 2006).

Second, recognizing a kludge on the basis of properties of cognitive or behavioral responses in which it is involved – as part of the responsible explanatory mechanism – still does not enable us to derive from these properties a specific algorithmic theory that can account for its formation. In section II.3.1.4 we discussed several strategies that could be employed for self-regulation, all involving a shift from controlled to automatic processing. This is clearly visible when comparing implementation intentions (Gollwitzer and Brandstatter 1997) with counterfactual thought (Galinsky and Moskowitz 2000): the first strategy primes or activates a particular response or components of it, whilst the second strategy does partly work through de-biasing or de-activating a particular representation instead. Still other strategies are possible as well, that do not so much focus on explicit and language-based representations, as aim to manipulate an agent's attention bias or behavioral approach tendency instead (Stacy and Wiers 2010). Moreover, in the previous section we noted that some reduction of the complexity of representations that are being processed can be obtained through their segmentation or chunking (Halford, Wilson et al. 1998). Consequently, a shift in processing could then occur. Common to all such strategies, however, is that they involve a form of self-regulation without constant conscious control. We are suggesting to consider this relatively stable adjustment of automatic processing as a result of kludge formation, such that the responsible mechanism yields more appropriate and self-regulated responses than before. It has to be acknowledged that the efficacy and durability of the established kludge can differ, since the mechanism yielding an agent's response can still operate with or without involvement of this kludge when activated in a particular situation.

Irrespective of the loose – not strict – independence of the neural implementation theory from the accounts of both the task at hand as well as its strategy that was defended in Part I, it is still relevant to consider the neural implementation of the kludge. However, given the multiple reasons given in section II.3.1.1 against the plausibility of a strict distinction between automatic and controlled processing, identifying neural correlates for the process of kludge formation may be complicated, too. Moreover, although there may be some agreement between authors about the nature and neural implementation of controlled processing – involving several PFC areas[220]-

[220] Involvement of several PFC areas for the representation of a cognitive or behavioral response is explained by the recognition that such a response relies on the representation of a distributed Structured Event Complex with multiple attributes that range from motor responses to social norms (Barbey, Krueger et al. 2009).

agreement is lacking regarding automatic processing (Evans 2008). Nonetheless, there are proposals regarding the neural correlates for the systems underlying automatic and for controlled processing respectively. Not surprisingly, a shift from controlled to automatic processing implies a diminishment of prefrontal lobe activity, which is recruited for controlled processing. The automatization of a previously controlled task likely depends upon increased basal ganglia activation, inter alia (Lieberman 2007).[221] Concurring with this is the account that focuses on the necessity to answer to the capacity limitations of working memory and executive control, equally implying that improved self-regulation without further challenging these limitations should diminish the reliance on PFC activation (Deutsch and Strack 2006). The latter account also refers to research of the basal ganglia as being involved in such a shift, while acknowledging that it is still unclear how and for what specific role the basal ganglia are involved. Processes as diverse as selection, inhibition, and attention direction have been mentioned as contributing to automatic executive control (Heyder, Suchan et al. 2004).[222] In sum, in terms of neural correlates, the kludge formation might be taken to consist largely of a shift from PFC activation to basal ganglia involvement, the latter being in need of further specification depending on the task at hand.

Our fourth kludge characteristic referred to the inter-individual variation visible in the kludge formation or in its final state. Such variation can depend from the large variation that is found between individuals in working memory, which will also affect when and how a shift in their modes processing occurs (Feldman Barrett, Tugade et al. 2004). Furthermore, as much as there are several forms of self-regulation or strategies that can support a shift in processing, these will add to the large variation between individuals or even between tasks. This variation is not only observable in task performances properties, but is also associated with differences in the mechanisms responsible for these performances. Indeed, Stanovich even rejects the notion of a general and task-independent automatic system but instead prefers to refer to TASS or 'The Autonomous Set of Systems', each of which can function as a distinct, autonomous system and can be triggered separately (Stanovich 2009). This emphasis on inter-

[221] Lieberman derives from his review of the literature on dual process theories the following neural correlates: automatic or reflexive processing is carried out with recruitment of amygdala, basal ganglia, ventromedial prefrontal cortex (VMPFC), lateral temporal cortex (LTC), and dorsal anterior cingulate cortex (dACC) and controlled or reflective processing recruits primarily lateral prefrontal cortex (LPFC), medial prefrontal cortex (MPFC), lateral parietal cortex (LPAC), medial parietal cortex (MPAC), medial temporal lobe (MTL), and rostral anterior cingulate cortex (rACC) (Lieberman 2007).

[222] A comparative account of the basal ganglia suggests that they should be considered a 'specialized device for the solution of selection problems' and can take over selection processes from the cortex, with which it is tightly connected (Redgrave, Prescott et al. 1999). Selection being an important and initial step of information processing, this account may at least explain part of the effect of automatization and expertise.

individual variability concurs with our earlier observation of different strategies that might be involved in a shift of processing, which will recruit more or less different neural systems as well.

Our fifth kludge characteristic refers to its employing pre-existing components, that might also be involved in other mechanisms. Similar to the fact that developmental learning was found to be largely a matter of modularization of processes that were already given (Karmiloff-Smith 1992), kludge formation in the present context equally depends on the 'neural re-use' that is prevalent in cognition and the brain (Anderson 2010). As noted earlier, particularly in section II.3.1.1, the assumption that two different processes can be distinguished does not imply that the two are strictly separate, nor that they rely on completely different systems. Indeed, when two different processes relying on different memory systems are apparently responsible for psychological and neuropsychological results, it is still useful to not strictly separate them but to recognize possible interactions between the two when modelling such results (Smith and DeCoster 2000). Even automatized reasoning tasks are shown to possibly invoke both processes, contrary to the belief that such forms of rationality can only be subserved by controlled processing (Stenning and Lambalgen 2008). More explicit in denying strict separation of the two types of processing is the uni-model account that assumes that a shift in processing merely involves the differential activation of a single model in response to specific parameters of the task at hand (Kruglanski and Orehek 2007) – suggesting a comprehensive explanatory mechanism which responds in a slightly modified manner to specific tasks, comparable to shifting gears or employing different components. Similar is the distinction of not two but four processes that interact differentially in task performances, in response to various task features (Sherman, Gawronski et al. 2008). In sum, dual-process theories do not assume that the kludge formation involved in shifting between processes depends upon the establishment or deployment of a novel mechanism.

Are kludges in this context involved in further dynamic trajectories, like being partly responsible for subsequent development or learning? Or are such kludges highly specific and only activated in very limited situations? Discussion of this sixth characteristic can be relatively short: insofar as kludges function as means to decrease processing demands, they allow subjects to perform increasingly complex and novel tasks in which such kludges are integrated. Indeed, automatization of a task via chunking helps to reduce the relational complexity of the relevant representations, allowing an agent to subsequently engage in even more complex tasks (Halford, Wilson et al. 1998). In the previous chapter and in the next part where the hierarchical structure involved in cognition and behavior is at stake, we argue that the performance

of increasingly complex tasks is dependent upon an agent's ability to form kludges, as when a shift to automatic processing of a particular task component occurs and allows him to shift his attention to another task component.

The final and seventh kludge characteristic pertains to the integration of environmental information. Not surprisingly, as with every developmental and learning process, environmental information is comprised in the eventual kludge. Indeed, evolution prefers open mechanisms precisely because they allow such including of environmental information, as can be observed once the imprinting mechanisms of chicks have integrated detailed information about their caregiver (Wimsatt 1986).[223] As kludge formation often involves an automatized informational process in response to specific environmental stimuli, it is plausible to assume that such information determines the kludge to a large extent. Interestingly, confirmation of this comes from a paradigm for targeted adjustment of an agent's kludge, which consists of his learning to automatically associate the specific stimulus with another response direction (Hofmann, Friese et al. 2008). Another self-regulating strategy influences kludge activation by using representations of a future task situation in which an intention should be implemented (Gollwitzer and Sheeran 2006), usually including both perceptual and linguistic information.

In the next chapter, we will more specifically consider how humans are particularly apt at kludge formation with the integration of environmental information, tools and language. In a certain sense it will complement our preceding arguments concerning strategies that aim to modify the representations involved in the relevant processes without calling upon other resources.

[223] It has even been argued that we should recognize that genes, too, contain information that is partly adapted to the environment and history of the organism (Collier 1998).