Left to right: Merlin Donald, Rob Kurzban, Michael Meyer-Hermann, Liz Spelke, Werner Güth, Christian Keysers, Jonathan Cohen, Eric Johnson, Jonathan Schooler, Julia Trommershäuser, Lael Schooler

11

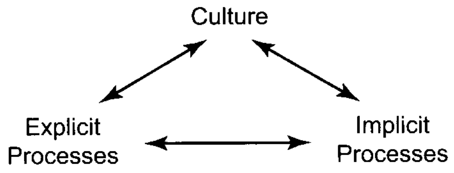
Explicit and Implicit Strategies in Decision Making

Christian Keysers, Rapporteur

Robert Boyd, Jonathan Cohen, Merlin Donald,
Werner Güth, Eric Johnson, Robert Kurzban, Lael J. Schooler,
Jonathan Schooler, Elizabeth Spelke, and Julia Trommershäuser

Introduction

Human decision making may be best understood as a triad. At the level of a single human individual, decision making depends on a variety of processes: some are more explicit whereas others have a more implicit nature. These two types of processes produce and are, in turn, influenced by, among other things, human culture.



As the scope of our discussion group was to examine human decision making, we begin with a discussion on the uniqueness of human cognition. Thereafter we explore the nature of explicit and implicit processes, and how they interact, and conclude by incorporating culture into decision making.

Uniquely Human

One way to hone in on an understanding of the relationship between implicit and explicit processes is to ask how animal cognition differs from human thought. When we consider the sensory and motor systems of animals and the brain systems that support them, humans look very similar to other animals

When we consider, however, the cognitive achievements of different species, we suddenly confront a chasm, for the products of human cognition far surpass those of any other animal. How should we intrepret this chasm? Does an appeal to the distinction between cognitive processes that are conscious versus unconscious, or implicit versus explicit, aid our understanding of the ever-increasing gap between humans and other species?

One attempt to address this question has focused on the capacities of human adults in relation to those of nonhuman animals. This research helps eliminate a number of intuitively plausible accounts of human uniqueness. For example, humans are not the only animals to use tools (as evidenced by the hooks of New Caledonian crows or termite fishing poles of chimpanzees; Hauser et al. 2002; Hunt et al. 2006) to communicate with symbols about aspects of the external world (e.g., the alarm calls of vervet monkeys; Cheney and Seyfarth 1990), or to reason about the epistemic states of others (as in gaze following and perhaps "theory of mind" reasoning in chimpanzees; Hare et al. 2001). Humans engage, however, in each of these activities far more extensively, systematically, and productively than any other animal. Along the timescale of evolution, the frequency and flexibility with which these capacities are used appear to show a sudden explosion in the transition from nonhuman primates to modern human beings (e.g., Mithen 1996), and this explosion needs to be explained. It is particularly visible in human infants and young children, who do not simply learn the specific tools, symbols, or patterns of behavior about which they get extensive information but who seek information from others actively so as to build encyclopedic knowledge about the world (e.g., a child who asks, "What is this for?" after every object in sight).

Accordingly, another set of attempts to address this question has focused on the capacities of human infants in relation to those of other animals. This research provides evidence that infants have systems from a very early age (in some cases at birth), for representing and reasoning about objects (Valenza et al. 2006), about other social agents (Farroni et al. 2002), and even about abstract entities in domains of geometry and number (Newcombe et al. 2000; Xu and Spelke 2000). For example, infants represent up to three inanimate objects exactly, they infer their motions and interactions in accord with basic physical constraints, and they maintain these representations when the objects move from view (e.g., Feigenson et al. 2004). As soon as children are able to locomote independently, they represent the distance and direction of their travel and make inferences about their changing position in accord with the rules of Euclidean geometry (e.g., Landau et al. 1981). In addition, infants represent the approximate cardinal value of large sets of elements, compare sets on the basis of cardinal value, and even perform operations on sets isomorphic to addition and subtraction (e.g., McCrink and Wynn 2004). In each of these cases, the cognitive systems found in infants continue to exist in older children and adults, and they support higher cognitive abilities such as physical, spatial, and numerical reasoning (e.g., Dehaene 1997; Scholl 2001). Nevertheless, nonhuman animals have been found to exhibit the same abilities, with the same signature patterns of performance (e.g., Hauser et al. 2003; for a review, see Spelke 2003a). These findings provide evidence that the cognitive systems underlying infants' achievements do not explain the species-specific cognitive capacities of humans.

Further attempts to illuminate human uniqueness focus on changes in cognitive capacities over human cognitive development. These studies provide evidence for both continuity and change in human cognition over development. On the one hand, the core cognitive systems found in human infants continue to exist and function throughout adulthood; they may also give rise to fundamental intuitions about the world as well as to rich systems of implicit knowledge. On the other hand, new systems emerge to bring about qualitative changes in human cognitive capacities that may underlie many of the hallmarks of mature, distinctively human cognition.

The development of number concepts serves as an example. As noted, infants have two systems for representing numbers: an exact system for distinguishing small numbers (1 vs. 2 vs. 3) and an approximate system for distinguishing large numbers with ratio precision (e.g., they discriminate 8 vs. 12 jellybeans as easily as 4 vs. 6). When children acquire the ability to count verbally, these two systems develop into one, which represents and operates over natural numbers. The stages of children's learning of counting are instructive (Wynn 1992): first they master the start of the list of number words (perhaps to "ten") and the meaning of "one." Then, in sequence, they learn the meanings of "two" and "three." Around four years of age, they induce that each word in the count list picks out a cardinal value that is one larger than the previous word. This induction appears to be beyond the grasp of the most highly trained animals (e.g., Matsuzawa 1985; Pepperberg 1987). In human children, this appears to result from systematically mapping the small exact number system, the large approximate number system, and the language of number words and verbal counting (Spelke 2003b).

In the case of numbers, therefore, uniquely human concepts and cognitive abilities result from the *productive combination* of concepts from preexisting "core" systems of representation that humans share with other animals. The same kind of developmental scenario appears to occur in other domains, including tool use (e.g., Bloom 2004), navigation (Spelke 2003b), and learning by observation. (Although nonhuman primates show rudiments of learning by observation and the mirror neurons that may be the prerequisite for imitation [Keysers and Perrett 2004; Keysers and Gazzola 2006], human adults can learn longer arbitrary sequences of actions far more rapidly and productively [e.g., Meltzoff and Prinz 2002].) Intuitively, it is appealing to suspect that this productive combinatorial capacity may relate to our capacity for explicit, conscious reasoning.

Language provides a particularly clear example of such productive, combinatorial explosion. Children first learn words, then very simple stereotyped phrases, and then suddenly the combinatorial power explodes and adults can make phrases of almost unlimited length. Interestingly, such phrases are not simple strings of words; they obey syntactic rules. The difference between simple sequences and syntax has been formally investigated along the distinction between sequences generated by Markov decision processes and recursive, phrase-structure grammars. A Markov decision process is one in which each new element in the sequence is generated by drawing it from a fixed probability distribution over the list of possibilities, independent of prior elements in the list. In this respect, Markov decision processes have no "memory." In recursive, phrase-structure grammars, however, elements show that long-distance dependencies and structured sequences can be embedded within other structures of the same type. Phrase structure and recursion are central to language and other domains, including number and theory of mind. This provides critical, additional functionality, including recursion; that is, the ability to embed structured "subsequences" within a sequence. Accordingly, this distinction is critical for understanding sequential behavior such as, in particular, motor planning and language. Human behavior in these domains demonstrates clearly the capacity for recursion, as in the sentence: "The cat the dog chased meowed." This seemingly simple capability provides tremendous additional representational power and may be at least one factor at the core of the explosive combinatorial capacity so characteristic of human behavior.

A variety of neural network architectures have been proposed that can learn finite state grammars, such as simple recurrent networks (e.g., Elman 1990) and gated architectures (Hochreiter and Schmidhuber 1997). In some cases, these have been related directly to neural mechanisms (e.g., Botvinick and Plaut 2004; Braver and Cohen 2000; Frank et al. 2001). However, a comparison of these mechanisms across primate species remains an important challenge for understanding what changes may have occurred in these mechanisms in the human brain, or what additional mechanisms are needed to support the tremendous flexibility of human thought and behavior.

Current experiments on decision making in monkeys could lead to fundamental insights into some of the basic steps that may form prerequisites for this explosion. When a monkey learns to shift his eyes to the right when seeing a rightward movement, and to the left following a leftward movement, decision making is performed in a rather limited perception–action loop. Monkeys, though, can be trained to view the moving pattern, store the results of this analysis for a while, and then report their percept by seeking a green or a red target, which can appear in any location of the screen. In such a case, the standard perception–action loop is interrupted, and what is activated during the perceptual decision making is not a motor program but a rule: "saccade to the red dot" or "saccade to the green dot." When the targets finally appear, the decision making for the motor system is no longer based on a direct percept, but may rely on prefrontal storage of more abstract information. This detachment may represent the first step along a continuum that makes it possible for humans to make decisions based on explicit rules and representations.

The issue of syntax and flexibility may be linked also to another preadaptation present in nonhuman primates: their motor behavior already exhibits some of the embedded structure typical of human language and other domains of cognition. Let us consider the following sequence of actions: "grasp a fruit and bring it to the mouth." First, this sequence has a structure that clearly goes beyond that of a Markov process: the probability of "bringing to the mouth" depends on the previous action, being much higher after grasping food than after most other actions. Second, action sequences can be recursive: if the food that is grasped is a banana, then the action of peeling the banana is inserted into the sequence, thus becoming, "grasp the fruit, peel it, and bring the peeled fruit to the mouth." In addition, just as words in a phrase can be replaced while preserving the other elements, objects in actions can be replaced while preserving the rest of the action: "grasp the flea from another monkey's fur and bring it to the mouth." To make the structure of actions available to more general domains of human cognition such as language, the domain specificity of actions must be overcome. If the neural architecture enabling the flexibility of actions only receives input from regions representing concrete objects and would only have output onto motor structures, it would be unlikely for these areas to be of use for language and thought. This does not, however, appear to be the case: the premotor and supplementary motor areas of the brain have widespread connections, including prefrontal, temporal, and parietal areas, which are involved in many domains of decision making (Rizzolatti and Luppino 2001). The idea behind the importance of the motor system for other domains of cognition comes from the fact that the left ventral premotor region involved in hand and mouth actions in monkeys (area F5) is considered to be the homologue of the posterior aspects of Broca's area (Brodmann area 44 in humans), lesions of which cause language impairments, in particular, of language production and grammar but also inner speech (Aziz-Zadeh et al. 2005). In addition, human subjects appear to use ventral premotor regions for a number of perceptual judgment tasks that are not inherently motor (Fiebach and Schubotz 2006). One possible explanation for this may be that during evolution, the remarkable capacity of premotor structures to organize actions into sequences may have been employed for many domains of cognition, thus giving these domains the sense of effortful, serial control that is typical for actions.

If the combinatorial and recursive abilities of human cognition are so central for modern humans, it may prove fruitful to investigate specifically the neural substrate for this capacity by contrasting it against the seemingly less flexible behavior that other primates exhibit.

One open question concerns the causal relationship between language and other flexible mental capacities in humans. It remains uncertain whether flexible language is a prerequisite for flexibility in the other domains. For example, some languages lack words to express exact numerical values beyond "two" or "three," and native speakers of these languages do not spontaneously represent exact numbers beyond those limits (Pica et al. 2004; Gordon 2004). These speakers' abilities to represent approximate numerosity and to perform approximate addition and subtraction on these representations are, however, indistinguishable from that of adults and children whose language and culture includes a full verbal number system, and who have learned symbolic arithmetic. It is unclear whether these speakers fail to develop precise number concepts beyond the limits of language because they have no words for these numbers or because their culture fails to teach them the meaning of these numbers, with the absence of words reflecting the lack of cultural knowledge about these higher numbers. Another example involves subjects who do not possess language, either because they were born deaf and were not taught a sign language, or as a result of brain injury. Here, individuals develop many of the unique human capacities, such as complex imitation and tool use, which suggests that language is not a prerequisite for all uniquely human capacities.

Through the evolution of language, the human brain has created a medium that facilitates the communication of knowledge in dramatic ways. The capacity for imitation coupled with explicit pedagogy might be important antecedents of human cultural products that are far greater than that of any other animal. This has created a dense network of communication between individuals that renders society a distributed network in which the decisions of each individual are influenced heavily by the other individuals in a process that we term "cultural."

The combinatorial and recursive capacity of human cognition, of course, is not the only aspect that differentiates humans from nonhuman primates. The capacity to plan deliberately by thinking about events in the past and future is far more developed in humans than in other primates. During fruit foraging, primates often deplete a fruit tree and return to the tree after a time sufficient for the tree to replenish its fruit. Given the complexity of their habitat, it is unlikely that there are adaptations designed for each particular plant; this would suggest that nonhuman primates may have at least rudimentary capacities to plan across time. Humans, however, can reason about, and routinely plan into, the distant future. The Inter-Temporal Choice paradigm, in which subjects are asked to choose between a small amount today and a larger amount after a time delay ranging from seconds to years, demonstrates that the decisions of subjects change gradually as a function of the delay, with a discount function that is far flatter than that of even apes. This suggests that humans clearly have an extended capacity to plan into the future.

Based on cross-species comparisons, Donald (1991, 1993, 1995, 2001) has proposed that the complex of behaviors and phenomena commonly described as "conscious processing" evolved gradually through a series of adaptations in various vertebrate species, each of which served to increase the degree of autonomous control over its own cognitive activities. Various aspects of conscious processing, such as alertness, selectivity, self-monitoring, and metacognitive oversight, may have evolved at different stages of evolution, for different adaptive reasons, and with different neural underpinnings. As a result, human conscious processing is enabled by a complex of neural structures, each of which has a unique function and evolutionary history. Human beings have a special kind of conscious control, distinguished by oversight, as shown in their exceptional human ability to self-rehearse skill and, especially, to automatize very complex skills that involve automatic serial attentional control. The primary adaptive purpose of conscious supervision of skills is not only to acquire skills, but to automatize them in a deliberate and consciously supervised manner. This capability emerged early in human evolution and was a precondition for the later evolution of language, and even of protolanguage (Donald 1998, 1999, 2005). It was also a precondition for the emergence of other kinds of advanced symbolic thought-skills, such as mathematics. Although this was not the only unique feature of conscious control in humans, it is perhaps the most characteristic one.

Overall, humans differ from nonhuman primates in a number of ways pertinent to decision making. They have a combinatorial capacity that is reflected in language, mathematics, and other domains. Their language and imitation capacities enable them to communicate more effectively than any other animal, and their tendency to teach each other deliberately may have enabled the development of a rich culture. These processes may have equipped humans with explicit mechanisms for coping with more complex cognitive demands than those of other animals, and these capacities have important implications for their decision-making processes.

Explicit—Implicit

Intuitively, human individuals believe that two types of processes occur in the mind. An example of the first process would be the act of tying shoe laces. Conscious control does not seem to be required, as one can, for example, talk about an unrelated topic while performing the task. In addition, it is difficult to verbalize the procedure. An example of the second process would be arithmetic calculation: 3 + 7 + 12 + 23. Here, conscious control is required, as evidenced by the degree to which performing another conscious operation (e.g., reciting the alphabet) interferes. Subjects can verbally describe parts of this process easily as well as carry out a continuous description of the operations; they also perceive that they are consciously aware of "what goes on in their heads" while performing the task. Nevertheless, research provides evidence that even the most explicit mathematical calculations are made possible, in part, through a host of automatic processes that are inaccessible to consciousness, including the automatic activation of core processes and approximate representations of numerical magnitude (e.g., Stanescu et al. 2000; for a review, see Dehaene 1997). When the core processes are impaired, mental arithmetic suffers greatly (e.g., Lemer et al. 2003). This illustrates that explicit cognitive processes cannot operate in isolation from implicit ones.

The *implicit association task* (Greenwald et al. 1998) is potentially interesting because it purports to assess the first type of process; namely, the association between, for example, a category and attitudes. In the canonical version, a subject is given a list of, say, African-American and European-American names as well as words with either good (happy) or bad (death) connotations. The task involves placing a word into a category depending on its content. In one condition, if the word is either the name of an African-American or something negative, one choice is indicated; if the word is either the name of a typical European-American or something positive, the other choice is selected. In another condition, if the word is the name of an African-American or positive, it is placed into one category, whereas if the word is either European-American or negative, it is placed into the other. Reaction times are measured and the difference in speed between the two tasks is taken to be a measure of the (implicit) association between the social category and negative (or positive) views. Reaction times are longer when African-American names are paired with positive words or European-American names are associated with negative words.

Subjects who show this implicit association will claim, however, to be unprejudiced, as measured by self-report scales (Greenwald et al. 1998). This is taken to be evidence that on one level—the implicit level—there is an association that is denied at another level—the explicit level—by the subjects. The implicit association task illustrates how psychologists have viewed and addressed empirically the implicit/explicit distinction.

It also reveals how a multisystem structure, with subsystems representing similar information in different forms (in this case, racial stereotypes), can be present in the brain as a whole, even though the subsystems conflict with one another (Kurzban and Aktipis 2007). To the extent that the computational mechanisms responsible for verbal report about the relevant questions do not have access to the information contained in the implicit associations that produce the experimental result, these associations do not influence verbal reports.

A number of different terminologies have been proposed for distinguishing these two types of processes (Table 11.1). Originating from different scientific traditions, these distinctions do not map perfectly onto each other. Defining subcategories of mental processes by several of these distinctions would suggest, for our purposes, a distinction that is too fine.

At this stage of research, it would be a vain enterprise to attempt to define unambiguously a set of features common to all processes placed either in the right- or left-hand columns of this table. Thus, to avoid bias in our discussions as well as in earlier versions of this report, we referred to processes as being those that fall within the left- or right-hand columns. This terminology, however, was felt to be cumbersome for the publication, and thus we refer to the processes as being reflexive or reflective.

Table 11.1 Separation of mental processes, originating from numerous scientific traditions, into two types of processes.

System 1 / Reflexive Processes	System 2 / Reflective Processes
Implicit	Explicit
Nonconscious	Conscious
Compiled	Interpreted
Procedural	Declarative
Automatic	Controlled
Habitual	Deliberate
Involuntary	Voluntary
Pre-preceptual	Perceptual
Pre-attentive	Attentive
Not registered	Remembered
Encapsulated	Accessible
Domain-specific	Domain-general

It is often felt that reflective processes are represented in a propositional format that can be transformed easily or explained in natural language, whereas reflexive processes are not. It is easy to explain how $(1 + 2) - (2 - 3)$ is equal to $1 + 2 - 2 + 3$, but it would be difficult for most of us to explain the motor process that allows us to run.

Behavioral performance in tasks that require reflective processes generally suffer if subjects are asked to perform attention and working memory-dependent tasks simultaneously, such as mental arithmetic (e.g., counting backwards in steps of 7 from 322), while reflexive processes are relatively less affected by such manipulations. Most activities require both and therefore suffer to some extent from interference from attention and working memory-dependent tasks. Consider the act of riding a bicycle: most people are able to ride while counting backward by threes, which shows that core skills involved in this activity might fall under reflexive processes. The speed at which you might ride, however, may be reduced by the dual task, and the risk of having an accident may increase, suggesting that skill can benefit from engaging reflective processes to complement reflexive processes.

Most people are generally aware of the source of information that governs reflective processes; however, they are often unaware of the sources of information governing reflexive processes. For instance, consider the following moral quandary:

> Christel and her brother John go to Europe. In the romantic atmosphere of Europe, they have protected sex. Do you think this is morally wrong?

Subjects generally answer yes, but when asked to verbalize the basis for their decision, most give reasons that are unlikely to be the true grounds for their decision (e.g., "because if they had a child, the child would be at risk of showing genetic deficiencies"). They might be unaware of key elements that drove their decision, which would point toward the existence of reflexive processes within their decisions (Haidt 2001). Alternatively, they might be truly aware of the motives behind their decision but fail to report these because they assume that a different rationale would be more desirable.

Our ultimate goal is to understand the neural mechanisms that underlie these two types of processes and the differences between them. Dimensions such as neural activity versus synaptic connections as well as local versus distributed representations are thought to be important (e.g., Cohen et al. 1996). For example, the flexibility of reflective processes, and their accessibility to report, may reflect their reliance on discrete patterns of neural activity used to represent critical components of the process (e.g., the intermediate products of a mental calculation or the subgoals of a sequence of actions). One feature of such representations may be the ability to activate them relatively independently of one another, and in arbitrary combination with others. This can be achieved in neural networks by local or functionally independent representations. This contrasts with distributed representations, which share overlapping feature sets. Distributed representations afford efficiency of coding (similar concepts share overlapping representations) and the ability for inference (activating a concept primes related ones). However, for the same reasons, they are less flexible. This distinction between local or functionally independent versus distributed representation may be an important factor underlying differences between both types of processes.

Another dimension may be the state versus weight distinction; that is, the difference between storing information in patterns of neural activity (states) versus changes in synaptic connection strengths (weights). Storing information as a pattern of activity in working memory (again, whether it be an item or a goal) is tremendously flexible (e.g., it can be modified quickly), but not very durable. Of course, information can also be stored in the nervous system through synaptic modification. Although in some cases this can happen quickly (e.g., in the hippocampus and amygdala), learning is thought to be more durable, but more gradual in the neocortex (e.g., McClelland et al. 1995). Such learning could support efficient processes (e.g., the execution of a routine sequence of actions), but would take time to learn and be less flexible. For example, when first learning to play the piano, a sequence of notes might be represented separately as distinct patterns of neural activity that activate other patterns of neural activity corresponding to the action needed to play that note. As a musical piece is learned, the entire sequence may come to be represented as a single pattern of neural activity as well as synaptic connections that allow the pattern of activity associated with each action to trigger the next action, thus bypassing representations of the notes themselves. The player, therefore, would have learned to replace an "explicit" or "declarative" representation of the notes (that may be consciously accessible) with a more "implicit" or "procedural" representation of the entire sequence. This would allow the actions to proceed quickly and efficiently from one to the next, presumably without direct involvement of the representation of each note by itself (and therefore perhaps less conscious awareness of, or access to, this information). This scenario corresponds closely to the perspective offered by some of the formal models described below.

Domain specificity may also be important for the distinctions. While reflexive processes are often specific for a particular cognitive domain, reflective processes might transcend the border of domains. Using the example above, representing the sequence as patterns of activity associated with each note, the person could play that sequence on any instrument for which they know the relevant actions. However, once this has been learned as a sequence of particular actions, it is specific to the instrument for which those actions are appropriate. This is similar to compiling a program for a particular machine. Written in a suitable language, the program can run on any machine that has a compiler for that language. However, once it has been compiled, it can only run on the type of machine for which it has been compiled.

The distinction between domain generality and domain specificity may also explain other phenomena. For example, because of their domain generality, reflective processes may have the peculiar characteristic of "disliking" contradictions. Thus, a series of propositions based on reflective processes (e.g., "I don't like my new neighbor," "My new neighbor is from the Congo," "Racism is bad") potentially creates cognitive dissonance that will lead to changes in some of these propositions in order to reach internal consistency. In contrast, because of their domain specificity, this is seen less often in reflexive processes: while observing the waterfall illusion, subjects report seeing movement but also report that individual features stay in the same location. Realizing this discrepancy creates a conscious sense of "oddness," which may be a reflective process, but does not lead to a change in the perception itself (reflexive processes), which appears to be immune to this dissonance. At the same time, when conflicting information occurs within the domain of reflexive processes, inconsistencies and instabilities can arise. Binocular rivalry and the Necker cube are good examples of this: incompatible sources of information produce unstable perceptual representations.

Reflective processes may also have the peculiarity of being deliberately triggered. For example, if someone nudges your back while you are standing, you will step forward to maintain your balance without feeling that you did so deliberately. If I ask you, "please step forward," you will take a similar action, but this time you will have the feeling that although I encouraged you to do so, you triggered the movement deliberately. The same action—stepping forward—can be accompanied by a feeling of automaticity or deliberate, self-triggering. This difference has some correspondence with neurology: patients with Parkinson's disease, for example, will step forward to maintain their balance, but they find it very difficult to step forward in a deliberate, self-triggered manner. A similar distinction can be applied to other modalities. Memories can be triggered externally (e.g., the smell of a Madeleine dipped in tea can trigger memories of the past) or self-triggered (the act of reviewing previous actions to recall where you left your keys). The latter process happens frequently in humans, but animals are also capable of deliberately conjuring visual memories (Naya et al. 1996).

Some of these distinctions depend on whether the process is accompanied by a feeling of "awareness"—a criterion that is both important and difficult. We feel intuitively that awareness is an important dimension of our psychological life, and referring to consciousness links research to the interest of the general public. Consciousness carries important implications for our legal system. A crime committed without realizing what one is doing is treated very differently to a similar crime committed without conscious understanding of one's act, for instance, under the influence of a drug (see Glimcher, this volume). Consciousness is, however, challenging to measure, as ultimately all appraisals must rely directly or indirectly on self-reports. The challenges of appraising consciousness can be readily seen by considering the different ways in which a person might be conscious of a cognitive event. An individual may be able to report that they experienced an event (e.g., the appearance of a recently presented face) without being able to describe the quality of that experience. Moreover, even the sheer judgment of whether or not the event was experienced may be subject to criterion factors (e.g., how conscious a signal must be for a person to characterize that experience as "conscious" depends on perceptions of the relative costs and benefits of a hit vs. a false alarm). Even if an individual is unambiguously experiencing a cognitive event (e.g., mind wandering while reading), it may take some time before that person notices explicitly that such events have taken place. Such subtleties in the measurement of consciousness raise important questions regarding the value of using consciousness per se as a heuristic for classifying processes, and thus is an area in which reasonable people can disagree. In our discussions, there was significant dissension regarding how pivotal the consciousness criterion should be in classifying cognitive operations. Some felt that consciousness itself is a red herring, and that the classification of cognitive states and processes should be based on more scientifically tractable constructs such as underlying neural substrates or computational processes. Others felt that despite its challenges, exploring the mapping between cognitive processes and consciousness is a worthwhile undertaking that can be enhanced through the careful recognition of the different types of conscious classification (e.g., detection vs. description, simply experienced vs. metacognitively appraised).

Although we disagreed on the degree to which consciousness should be viewed as pivotal in the classification of cognitive processes in Table 11.1, we agreed that two approaches seemed promising. First, systematically examining in which respects the different distinctions proposed in Table 11.1 overlap may help us explore the core properties of what we believe to be the two main clusters of processes. Second, the identification of particular examples of processes that systematically fall within the right- or the left-hand column of Table 11.1, independent of terminology, could provide prototypical examples of these processes. These examples could then be studied using evolutionary, neuroscientific, and computational approaches to provide snapshots of how our brain implements these different processes and why, in evolutionary terms, we have such mechanisms. The resultant understanding could be used to form the basis of a better understanding of the clusters they represent.

An additional approach is to map this distinction—whether categorical or representational of an underlying continuum—onto underlying neural mechanisms. In doing so, however, we need to exercise caution. Although the distinctions in Table 11.1 have received considerable support in the behavioral sciences, they may not map neatly onto neural mechanisms or architectures. In particular, it seems likely that performance of any *task* will involve a mix of *processes* from each column. It may therefore be useful to distinguish between tasks and processes. For example, even the most commonly used examples of automatic *tasks* (e.g., driving a car), which involve mostly unconscious or procedural *processes*, typically require an explicit and conscious intent to initiate the task. Conversely, tasks commonly used to highlight reflective processes (e.g., novice performance on a musical instrument) involve many automatic, unconscious, and procedural processes (e.g., the actions associated with playing each note, each of which is itself composed of a complex sequence of precise motor commands). Even the earliest efforts to identify the neural mechanisms underlying both types of processes have revealed a high degree of overlap, such as the role of dopamine in simple conditioning (Montague et al. 1996) and its role in gating of representations into working memory (Braver and Cohen 2000; Frank et al. 2001).

Even if the relationships in Table 11.1 and the underlying neural mechanisms are complex, this does not mean that the distinctions in Table 11.1 are useless. For some purposes, at other levels of analysis, these distinctions may reflect useful functional abstractions that serve as valuable heuristics for understanding higher-level phenomena (e.g., social interactions). Accordingly, the approaches we have described are not limited to decision making: in the understanding of social cognition, progress has been made even without clear definitions of the object of study. The need for terms such as explicit and implicit may actually disappear as a better understanding of the processes emerges.

Transfer between Processes

Interestingly, during behavior, human beings routinely change how strongly they rely on reflexive and reflective processes. One example in representational format is the learning of a motor skill. As described above, at first, representations about playing a piano might be in propositional form: "play middle C with the index finger of your right hand." The novice might experience playing the piano in the early stages as the feeling of executing these sorts of propositions and may indeed interpret notes in this way, as propositions directing their actions. Later, with practice and repetition, an expert piano player's movements will be caused by procedural representations, not in propositional form, but rather in the format of procedural memory. A reasonable characterization of the causal etiology of a novice playing the piano is a set of propositional representations, whereas the causal etiology of the muscle movements in the expert is a set of representations in procedural formats (Squire 1987).

In sum, it is fruitless to ask whether reflexive and reflective processes generally lead to better performance. It would be better to ask what mix of processes is involved in different circumstances. A number of approaches have allowed us to examine the contribution of the two types of processes in different situations.

Neuroscience

Functional neuroimaging allows us to examine the brain circuits involved in a number of tasks which, in turn, can be helpful in identifying the processes involved. Consider the footbridge scenario (Cohen 2005):

> As a bystander on a footbridge that crosses over a track, you see an oncoming trolley that will most certainly kill five individuals working down the track. You must decide whether to push a large worker, who is standing on the bridge close to its edge, off the bridge to block the oncoming trolley. You realize that this will certainly kill the large worker but that the action will spare the five workers further down the track.

At least two streams of information processes would be activated in this process: one is similar to that observed when certain emotional processes are engaged; the other is commonly associated with more deliberative forms of processing (Greene and Haidt 2002; Cohen 2005). Similar dissociations can be found when subjects have to decide whether they prefer to receive $10 today or $12 in a week's time (inter-temporal choice). Generally, these findings have been taken as evidence that many decision problems are influenced by both controlled, deliberative, reflective processes and emotional, less-controlled mechanisms that might better be classified as reflexive processes.

What can we deduce from fMRI experiments? By measuring the disruptive effect of deoxyhaemoglobin on the spin of protons in the brain, fMRI can indirectly localize changes of metabolic demands in the brain. The spatial resolution of the method is higher than that of many other neuroimaging techniques (e.g., EEG, MEG and TMS) but remains coarse compared to the size of the neurons and synapses that actually process information in the brain: fMRI

typically measures activity in spatial units of 3 mm × 3 mm × 3 mm = 27 mm^3. Within this volume, the method averages the metabolic demands of about 2500 million synapses and 25 million neurons organized in approximately 10 distinct functional columns. This spatial limitation constrains the interpretation of fMRI findings substantially. If one finds, for instance, that lowering the amount offered in an ultimatum game results in the recruitment of additional brain areas (Sanfey et al. 2003), it is reasonable to conclude that additional *processes* have been recruited by the new task parameters. What is difficult to do with fMRI, on the other hand, is to gain information about the *nature* of the additional processes that have been recruited: looking *where* in the brain the task causes activity cannot tell us directly *what* happens in that location of the brain. Many researchers interpret activation of a particular brain region "A" (e.g., the anterior insula) by referring to evidence that the same brain region was recruited in a number of other tasks, all of which are thought to share a certain process X (e.g., processing aversive emotional material; Wicker et al. 2003). This evidence is then used to suggest, for example, that brain activity in the insula during the ultimatum game indicates that the task involves the processing of one's negative emotional responses during unfair offers. This line of thought is tempting, because rejecting unfair offers often goes hand in hand with a negative emotional feeling, and the anterior insula is indeed active while rejecting ultimatum offers (Sanfey et al. 2003) and while subjects smell disgusting odors (Wicker et al. 2003). This can be taken to suggest that processing one's own feeling of disgust *might* be an element in rejecting unfair offers. It is, however, important to realize that areas such as the anterior insula contain hundreds of millions of neurons, and finding that such an area is recruited during the experience of disgust is not evidence that all of these neurons participate exclusively in the experience of disgust. Indeed, the anterior insula is involved in processing a range of emotions, including pain and positive emotions (Jabbi et al. 2007) as well as disgust and a variety of more cognitive tasks involving language. With that caveat in mind, finding that a task activates the anterior insula, which has often been found to be involved in emotional processes, *suggests* that the task requires emotional processing, but it does not *prove* that it involves such a process. At present, for many brain areas, our understanding of their function remains too scant to allow us to conclude *what* processes are involved in a task based on the location of activation. This interpretative limitation is common to all neuroimaging studies, and there is an awareness of this throughout the neuroscientific community. At the border between disciplines, however, neuroimaging studies are increasingly referred to by scholars outside the field of neuroscience. It is thus important for these scholars to be aware of the tentative nature of the interpretation of neuroimaging data.

Modeling

The development of formal models of decision making and their implementation in artificial systems has provided a fruitful testing ground for our understanding of decision making. In particular, it has helped identify the advantages and disadvantages of the processes shown in Table 11.1.

One fundamental trade-off between processes is the efficiency but rigidity of reflexive processes versus the flexibility but inefficiency of reflective processes. This distinction has been captured by several formalisms: model-free versus model-based learning mechanisms (Daw et al. 2007). Standard reinforcement learning algorithms provide an example of a model-free mechanism (sometimes referred to as "habit learning"). These models learn to represent the future value of each state in an environment. With sufficient time and stability of the environment, reinforcement learning algorithms can be shown to converge on an accurate representation of the expected future value of every state based on the future rewards with which each is associated and their probability of occurrence. However, learning the state-value associations is slow (requiring an exploration of all states in the environment), and if the environment changes (e.g., the probabilities of states changes), then the value representations can quickly become obsolete. In this sense, the system is inflexible—it has a hard time adapting to new environments. This contrasts with a mechanism that has a reflective-type representation (i.e., a model) of all perceived states in the world (and their associated rewards) and can simulate the various sequences of actions that can be taken from a given state to determine its future value. This system is flexible, inasmuch as it can compute the value of any state "on the fly." It is thus robust to changes in the environment, since the model can be quickly updated when the world changes. However, its ability to carry out simulations is limited both by its memory capacity (i.e., the length of sequences that it can simulate) and the time it takes to carry out such simulations, and thus it is less efficient than a model-free system. This trade-off between efficiency and flexibility seems to capture an important distinction between both types of processes.

ACT-R theory provides another formalism within which to describe this trade-off between efficiency and flexibility in reflexive versus reflective processes. In ACT-R this can be explained by the relative degree to which a process relies on each of two fundamental memory mechanisms: declarative and procedural memory. The core of ACT-R is constituted by the declarative memory system for facts (knowing what) and the procedural system for rules (knowing how). The declarative memory system consists of records that represent information, for example, about the outside world, about oneself, or about possible actions. The procedural system consists of if-then rules that guide the course of actions an individual takes when performing a specific task. The "if" side of a production rule specifies various conditions: the state of working memory, changes in perceptual information (e.g., detecting that a new object has appeared). If all conditions of a production rule are met, then the rule fires and the actions specified (the "then" side of the rule) are executed. These actions can include updating records, creating new records, setting goals, and initiating motor responses.

To see how processing in ACT-R relates to the distinction between model-free and model-based learning mechanisms, consider the situation of trying to guess your friend's two-letter pin code to access his bank account. Initially, you would explore the environment, trying out different pin codes. This could be accomplished by trying out various sequences: if want-pin then try "3" as first number; if want-pin then try "8" as second number, etc. This yields a highly flexible system that would eventually find the pin, say, "13" and produce a production of the form: if want-pin, then try "13." Over time, as you consistently gain access to your friend's account with this sequence, guided by a process of reinforcement learning, the "13" production would be strengthened, increasing the likelihood of firing and the speed with which it fires. Thus, it would win out over the simpler but more flexible productions that lead to exploration. However, this would lead to a problem if your friend changes his pin code. Here, the system would do well to return to the more flexible, but slower model-based production that can explore the environment. Notice, however, that in the ACT-R implementation, there would not be two independent systems. Instead, in reflective processes, processing relies more heavily on the declarative system (corresponding to the representation of the model in model-based learning), whereas in reflexive processes there is greater reliance on the procedural system. The general point is that there are trade-offs between more automatic, ballistic (i.e., procedural, in ACT-R) processes and more controlled, flexible ones that rely more on declarative memory. Which works best depends on the structure, stability, and experience with the environment.

If we include the reasonable assumption that we are consciously aware of changes to the goal representation that guides our behavior, we would end up with a situation where we are conscious of fishing around for the pin code by using explorative production, but unconscious of typing the pin when using the chunked "13" production. Consciousness—a distinction that has been used to distinguish system 1 and system 2—in the ACT-R implementation falls out of a single set of mechanisms, but understanding how this translates to the human brain remains uncertain.

Behavior

As discussed above, infants have approximate concepts of number that continue in adulthood. In addition, however, adults have precise number concepts. Adult number processes suffer from attentional load and would thus be classified as reflective processes, whereas the former do not. Both number processes could be used to investigate the dual nature and interaction of decision making. For instance, many models of decision making suggest that quantities are integrated *as if* they were processed using an explicit arithmetic strategy. Outcomes, for example, are weighted by their probabilities, or amounts are discounted by delay time. There are many reasons to suggest that various decisions are not made using an explicit arithmetic strategy: Protocol analysis shows that such explicit multiplication is rare, and analysis of reaction times does not seem consistent with explicit multiplication. In addition, analyses of information acquisition show that people seem to look at probabilities and payoffs contiguously, which seems inconsistent with the use of explicit arithmetic strategies. This suggests that behavioral experiments can be used to examine the degree to which decisions follow the psychometric laws of reflective processes, explicit arithmetic of reflexive processes, and approximate numerosity. An example might be to measure the subject's psychometric sensitivities along their intuitive sense of numerosity and to examine whether their choice behavior in economic decisions is better predicted by the characteristics of their approximate or explicit mathematical capacities, or a weighted combination of both. Placing subjects under constraints of high attentional load or allowing them only a brief instant to make decisions may be another way to probe the role of approximate numerosities in decision making.

The Impact of Making Implicit Knowledge Explicit

One strategy to study the relationship between implicit and explicit knowledge is to ask: What is the impact of making implicit knowledge explicit? The core intuition that can be affected by the use of implicit knowledge to make it explicit is illustrated using the example of motor processes. Everyone has had experience with trying to decompose a well-learned motor skill (e.g., explaining how to perform a tennis serve) only to find that in the attempt to characterize the process, it loses its automatic fluency. The basic claim that implicit processes can be affected (often negatively) by explication has been empirically illustrated in a variety of different domains, including memory, problem solving, and decision making.

Memory

Numerous studies have shown that attempting to describe a nonverbal stimulus can impair subsequent recognition performance. For example, Schooler and Engstler-Schooler (1990) showed participants in a study a video of a bank robbery. Some participants were subsequently asked to describe the appearance of the face of the robber in detail, while others were given an unrelated filler activity. Finally, all participants were given a forced choice recognition test, including a photo of the previously seen robber and seven similar distractor faces. The results revealed that describing the face led to a markedly impaired recognition performance relative to those who did not describe the face. Since the original demonstration of this negative consequence of explicit verbalization (termed verbal overshadowing), the approach can be generalized to other types of nonverbal stimuli, including memory for colors (Schooler and Engstler-Schooler 1990), music (Houser, Fiore, and Schooler, unpublished), voices (Perfect et al. 2002), abstract figures (Brandimonte et al. 1997), wines (Melcher and Schooler 1996), and mushrooms (Melcher and Schooler 2004).

One critical issue in understanding the impact of experiences that are not able to be explained involves the question of how expertise mediates this effect. A reasonable hypothesis is that when individuals are first acquiring a skill, explicit verbal description should be at a minimum benign and quite possibly helpful. However, as the skill becomes proceduralized, it may become more vulnerable to verbal explication. At the same time, if in the process of gaining expertise individuals also gain skill in verbally describing their knowledge, verbalization might not be disruptive. A study on the effects of verbalization on memory for the tastes of wine illustrates these points. Melcher and Schooler (1996) asked participants to taste a wine and either verbalize it or perform a control task prior to identifying the tasted wine among distracters. Participants were non-wine drinkers, untrained wine drinkers, or trained wine experts (i.e., professionals or those who had taken wine seminars). This study showed that verbalization only impaired untrained wine drinkers. Trained wine drinkers have the verbal ability to describe the various aspects of wine detected by their palates and do not experience verbal overshadowing, ostensibly because their experiences and verbalizations correspond. Similarly, non-wine drinkers have neither the expertise to perceive wines in depth nor the verbal tools to describe them, and thus their experiences and verbalizations do not interfere with each other. Untrained wine drinkers' perceptual expertise in wine tasting exceeds their verbal expertise in describing wine, and thus they do not have the verbal tools to describe all of the nuances that their palates detected.

Problem Solving

Over the years, various researchers have speculated that the processes leading to solutions on insight problems, in which the solution seems to pop out of the blue, are markedly less explicitly reportable relative to more logical or analytical problem-solving tasks. Accordingly, if verbal explication of implicit processes can interfere with those processes, then verbal description of insight problem solving might be disruptive. Consistent with this prediction, Schooler et al. (1993) found that participants who were asked to think aloud while trying to solve insight problems were significantly less successful than participants who worked on the problems silently. In contrast, thinking aloud had no effect on participants' ability to solve logical problems. Furthermore, in a protocol analysis, Melcher and Schooler (1996) found that the contents of participants' verbal transcripts were predictive of their success at solving logical problems, but not insight problems. This suggests that the source of the disruptive effects of explicating insight problem solving stems from the fact that the processes contributing to successful solutions are not available for verbal report.

Decision Making

A number of studies have demonstrated that verbal reflection can disrupt the quality of affective decisions. In one study, participants were given the opportunity to taste various strawberry jams (Wilson and Schooler 1991). In one condition, participants analyzed why they felt the way they did about the jams, whereas in another they did not. Thereafter, all participants rated the jams. When compared with the ratings of jam experts (who were from the Consumers Union), participants who went simply with their intuitions were in marked agreement with the experts. However, participants who analyzed their reasons generated evaluations that did not correspond with expert opinion. A follow-up study examined the impact of analyzing reasons on post-choice satisfaction. Participants were shown various art posters and were allowed to select one to take home. Prior to their decision, half of the participants analyzed why they felt the way they did about the posters. When contacted two weeks later, these individuals were less satisfied with their choices and less likely to have the posters hung on their walls. Such findings suggest that, as in the case of memory and problem solving, explicit reflection about decisions that typically draw on nonreportable processes can disrupt the quality of those decisions.

Summary

In considering the evidence for disruptive effects of explicit reflection on tasks that draw on implicit knowledge, several critical issues arise. First, what are the underlying mechanisms driving these effects? Two general classes of mechanisms have been proposed. According to a specific *content account*, verbal reflection is disruptive because in the course of describing an experience, individuals generate explicit verbal content that leads them astray. For example, individuals might erroneously describe the appearance of a face and then choose one that best fits their flawed description. According to a general *processing account*, verbal reflection leads to a generalized shift in the basic cognitive operations in which participants engage. For example, verbalization could cause participants to shift from an automatic holistic face recognition strategy (e.g., choosing the face that "pops out") to a more strategic analytic strategy (e.g., systematically comparing the faces feature by feature). At present, there is debate regarding which of these accounts best characterizes the findings, although it is notable that there are some results that are very difficult to accommodate with a processing account. For example, if, following verbal reflection, individuals rely on the specific content of what they said, then it is not self-evident why describing one face would interfere with the recognition of a different face that had not been described.

Importantly, the existence of suggestive (if not conclusive) evidence supporting a processing account of the disruptive effects of verbalization in the case of face recognition does not rule out the possibility that verbal overshadowing effects in other domains might not be more driven by the specific content of the verbalizations. This observation highlights a more general point, which is that there are a variety of notable differences among the domains in which the negative effects of explicit reflection have been observed. In the area of memory, the qualities that make stimuli difficult to describe most likely stem from their perceptual nature. It is not that the experience of a face or a taste is unconscious, it is simply that such experiences do not readily lend themselves to translation into words. In contrast, in the domain of insight problem solving, the nonreportable processes that lead to a solution may genuinely be unconscious in nature. Finally, in the case of decision making, the source of the disruption of verbal reflection may stem not from the perceptual nature of the stimuli but rather from the fact that the decisions are being based on an affective intuitive reaction to the stimuli that may simply not lend itself to reflective penetration. The reasons behind the lack of capacity for explicit reporting may vary markedly, highlighting the main point; namely, that implicit knowledge does not necessarily originate from the same underlying cognitive mechanisms.

Although caution is advised in considering the shared mechanism and underlying representations leading to verbal overshadowing effects across domains, there are some commonalities. In the memory domain, verbal overshadowing effects are attenuated if, following verbal description, participants are forced to make their recognition decisions quickly. Similarly, in the domain of affective decision making, if individuals are forced to make snap decisions, the negative consequences of previously having reflected verbally on the basis of their preferences can be attenuated. This suggests that despite the disparities between the domains, forcing individuals to reapply implicit processes negates the negative consequences of making implicit explicit.

Too Bad—Two is Bad: Problems with a Dual Process Perspective

Many lines of thought converge to the idea that humans have two distinguishable *types* of processes that occur in the brain and that these two types are not mutually exclusive. The process of decision making, however, can also be examined at an even more global or local perspective.

Two Is Too Many: Two Horses but a Single Cart

When adopting the perspective of the organism as a whole, it becomes obvious that the organism has to settle ultimately for a single decision at a given moment in time. In a moral dilemma, as in the footbridge scenario, one has to decide whether to push the large person from the bridge in order to save five individuals, or to refrain and let the five individuals die. If one assumes that a more emotional process would tend toward not pushing the person over the bridge, while a more deliberate calculating process would favor this act, the question remains as to how these two processes are integrated to reach a final decision. A similar line of reasoning may apply to the decision of whether to accept $10 today or to wait and receive $12 in a week's time.

The two proposed processes may perform entirely distinct computations and compete in an unsupervised fashion against each other. This may be the case if both processes feed onto the same output node, with one sending more excitatory inputs toward nodes that would push the large individual, while the other sends more inhibitory inputs toward these nodes. The two processes may also have direct inhibitory connections between each other. In such a case, it could be argued that the system will ultimately reach a single conclusion and that relatively little integration occurs between the processes.

Viewed differently, the two sets of processes calculate their own decisions but, in addition, another process is added before access is granted to motor output. This additional decision process could decide, based on the particular situation, how to weigh the decision of the two processes. For instance, a trained magistrate may have learned not to trust his emotions when making moral decisions and may thus base his decision entirely on the outcome of the deliberate calculating process ("one death is better than five"). Alternatively, after reading scientific publications, the same magistrate may decide to trust his gut feeling when deciding what painting to buy for his living room. Under such a perspective, the system may be seen as less competitive and more as an integrated whole that draws on multiple subsystems. A clear challenge for such an architecture is that the superordinate decision process would need to optimize some form of general utility function. Some have interpreted existing data as suggesting that humans do not maximize a single, simple utility function, which in turn would imply that it is unlikely that the brain routinely implements fully rational superordinate decisions (see, e.g., Camerer 2003). Perhaps individual computational processes generate outputs—such as a decision—as well as a probability that the decision is the correct one (on some standard).

Overall, it should be noted that the idea of integrated decision making does not mean that the brain implements a single uniform decision process per se. The brain is a mosaic of regions with different patterns of input and output and distinct cytoarchitectonics. Brain regions, therefore, differ in the types of computations they can perform. These unique processes are then integrated through an intricate network of synaptic connections between brain areas.

A potential mechanism through which different processing streams of decision making in the brain could integrate with each other is linked to the fact that most brain areas have reciprocal connections with each other. If area A and area B both provide independent decisions and feed these decisions to a subsequent area C, then the influence of area A on the decision in C will be sent back not only to area A but also area B; the same applies to the influence of area B on the decision in C. A single stable decision state in the system can then be achieved through the forward and backward propagation of evidence for a particular decision, providing more integration than would be assumed without backwards connections.

Two Is Too Few!

Table 11.1 conveys a set of distinctions that are common and long-standing in the behavioral science literature. Although these broad distinctions may be useful, it is worth questioning just how similar the mechanisms are that underlie the processes of each type. With regard to reflexive processes, there is general consensus that different processes must engage distinct mechanisms. That is, specific instances of automatic or implicit processes (e.g., driving a car, playing a musical instrument, or responding emotionally to a movie scene) certainly involve very different mechanisms, even if they share commonalities at some level. This is consistent with the high degree of domain specificity of such processes, as discussed earlier. Detailed mathematical and computational models are beginning to emerge that characterize formally the mechanisms involved in such processes (e.g., retrieval from episodic memory, semantic priming, sequential action). This should help illuminate both the differences and commonalities among such processes (e.g., Botvinick and Plaut 2004; McClelland, McNaughton and O'Reilly 1995; Ratcliff and McKoon 1988, 1996, 1997).

There is less consensus about reflective processes. Traditional cognitive psychological theory postulated that these share a central attentional resource (e.g., Posner and Snyder 1975; Shiffrin and Schneider 1977; Baddeley 1981), as evidenced by their ability to interfere with each other (for dissenting arguments, see Allport 1982; Navon and Gopher 1979). This would suggest that reflective processes rely on a common set of mechanisms, such as a single attentional and working memory mechanism with limited resources. However, more recent work suggests that there may be greater diversity than had previously been thought, including evidence for domain specificity in working memory processes (e.g., Shah and Miyake 1996) and the operation of different types of control mechanisms (e.g., Cohen and O'Reilly 1996; Braver et al. 2003). Thus, it seems likely that both types of processes have diverse mechanisms that share a common but not identical set of properties. Again, the specification of formal models in conjunction with a search for neural mechanisms will help refine our understanding of the relationship between and within these categories.

Summary

Although it may be helpful for current research to use the heuristic that there might be two *types* of processes contributing to decision making, we feel that it is imperative to collect further empirical evidence through a multidisciplinary approach. Such investigations will need to integrate the neurosciences, the study of behavior, and computational modeling. The resulting empirical evidence will focus our thinking and hopefully permit an understanding of the formal structure of decision making to emerge.

Culture

Individual decision making appears to be composed of multiple processes, yet the complexity of human decision making does not stop with the individual. The human brain and the decisions that humans have made have shaped our environment and our culture dramatically. The unique properties of human cognition outlined above enable us to talk to each other in high-rise buildings and to generate reports. No other animal has produced a culture of that sophistication or shaped its environment as much as we have.

One way to examine the effect of our own decisions on culture is to view culture as a result of individuals linked together through language, imitation, and learning by observation. This link transforms us as individuals into a network of brains that is more complex than the simple sum of its parts (Donald, this volume) and represents one of the major evolutionary cognitive innovations of hominids. Human beings perceive, remember, and make decisions in groups. They also trade memories and ideas, and build systems and technologies that regulate, and sometimes radically modify, the nature and patterning of cognitive activity in groups. Human culture has thus evolved into a storehouse of knowledge and procedures that are collectively available.

Human adult brains are the result of an intimate interaction between brains and culture, and the human brain is specifically adapted for connecting to cognitive–cultural networks (CCNs). Many of the capabilities that comprise "higher cognition," including language, do not appear in development without an efficient connection to CCNs. Some of the skills needed for building an effective connection with culture (e.g., joint attention, theory of mind, and perspective taking) are themselves dependent on cultural input. In that sense, the human brain cannot realize its design potential without extensive cultural input.

Culture Influences Decision Making

The relationship between brain and culture is bidirectional. Obviously, the brain influences culture. Canine brains produce canine cultures, primate brains primate cultures, and human brains human cultures. For humans, however, the reverse influence is much greater than in other species: the human brain receives from culture a great deal of specific formatting that enables it to use symbols. Moreover, the symbols themselves are conventions generated in cultural networks.

The clearest and best-documented case of culture influencing the brain is literacy. Writing is relatively new in human history and still far from universal in human society. Since the vast majority of human languages have no indigenous writing system, a specific brain adaptation for literacy could not have evolved. Literacy skill is a cultural innovation, imposed by culture through prolonged education, and yet it results in a quasi-modular brain architecture (Donald 1991). The neuropsychology of literacy has been well documented in cases of acquired dyslexia, and the brain systems that support literacy skill can break down in a double-dissociated manner, with patients selectively losing only part of the system. Thus, one can lose the ability to read irregular words independently of the ability to read regular words, and vice versa (Shallice 1988). Brain imaging studies further demonstrate the impact of literacy on brain organization (Petersson et al. 2000; Li et al. 2006; Stewart et al. 2003), in particular by showing that a visual area in the left occipito-temporal sulcus becomes partially specialized for visual word recognition in a given script during reading acquisition (Baker et al. 2007; McCandliss et al. 2003; Vinckier et al. 2007).

From this example alone, one can conclude that culture is able to affect significantly the functional architecture of the nervous system. Note, however, that this influence stays within bounds. Comparative fMRI studies indicate that a universal brain network is activated in readers of writing systems as different as English, French, Hebrew, or Chinese, with minor differences as a function of script (Bolger et al. 2005; Nakamura et al. 2005; Paulesu et al. 2001). Dehaene (2007) argues that the invention of writing systems was tightly constrained by the prior functional architecture of the brain and thus the interaction between brain structures and cultural inventions is bidirectional: cerebral networks can be partially "recycled" to a novel cultural use, but they also impose important constraints on which cultural inventions are successful (Dehaene 2007; for a similar argument concerning mathematical invention, see Dehaene 1997).

Developmental theorists have proposed that culture provides essential input in the development of such fundamental cognitive capacities as joint attention, word-finding, and social cognition (Nelson 1996; Tomasello 1990). These capacities are highly evolved in humans, but can only develop normally with extensive cultural guidance. Because culture influences the functional organization of the cognitive system so deeply in ontogenesis, *ipso facto* it affects the apparatus of decision making.

Inherent in the idea that human cognition is closely interdependent on CCNs is the realization that our brains and decisions are also heavily influenced by our culture in very specific contexts. Over the last years, this effect has been extensively investigated.

One example of how explicit cultural concepts can influence individual decision making is illustrated by examining the relationship between beliefs about free will and moral action. In a recent study (Vohs and Schooler 2007), participants were given an excerpt drawn from Francis Crick's, "*The Astonishing Hypothesis,*" in which he argues that humans are simply machines and lack any real sense of free will. Other participants were given a comparably long section of text that made no reference to the issue of free will. All participants were then given a mental arithmetic task for which there was an opportunity to cheat. Vohs and Schooler found that participants who read the article were significantly more likely to cheat, and that this increased cheating behavior was mediated by a reduced belief in free will. In another study, participants who were given deterministic statements were more likely to overpay themselves for their performance on a problem-solving task. Together, these studies illustrate how establishing certain assumptions explicitly about human nature can impact powerfully, and in some cases perhaps detrimentally, people's moral decisions.

One can also compare decision-making behavior in different cultures. Recently, a team of anthropologists and economists performed two rounds of experimental games in a wide range of small-scale societies, including foragers, small-scale horticulturalists, pastoralists, and sedentary farmers (Henrich 2000; Henrich et al. 2004, 2005). One of the tasks was the ultimatum game, which allows two parties (the proposer and the responder) to share a gross positive amount of money. The proposer suggests a distribution of this amount among the two, which the responder can either accept or reject. Acceptance results in the proposed distribution, whereas rejection leaves both parties empty-handed. Another task was the dictator game, where there is no veto power (i.e., whatever the proposer suggests is automatically implemented). The third type of game was the so-called public goods game. Here each of several players can either invest a part of their capital into the "public good" or keep all of their money. At the end of each round, a player earned what he kept plus what he got from the sum of all contributions after the contributions had been multiplied by a certain factor determining the amount of the public good. Parameters are such that individually it pays to keep everything (i.e., contribute nothing), although the group of all players would gain by contributing. Under certain conditions, third-party punishment was implemented by letting a third party watch how the money amount is allocated in the dictator game. Being aware of the distribution, the third party had the choice of subtracting a gross positive punishment from what the "dictator" took. In case of punishing, this subtracted amount was lost (i.e., it was neither given to the recipient nor collected by the third party).

In the first round, the ultimatum game was performed in all societies, whereas the public goods game and the dictator game were performed only in a subset of societies. In the second round, ultimatum, dictator, and third-party punishment games were performed in all societies. These experiments reveal a number of interesting results:

1. *Behavior in the ultimatum game varies widely.* Modal offers (i.e., how much proposers assign to responders) vary from above 50% to as low as 15%, and in some societies such low offers were usually accepted. This behavior is closer to predictions based on opportunism in the sense of own money maximization than the behavior of Western university subjects.

2. *Behavioral differences are correlated with group characteristics but not individual characteristics.* Individual ultimatum game offers were not significantly correlated with individual characteristics (e.g., income, wealth, education, market contact, age or sex). However, multiple linear regressions showed that more average aggregate market contact and cooperation in subsistence increases significantly average ultimatum game offers, and together the two variables accounted for more than half of the variance in average offers among groups.
3. *Variation in punishment predicts variation in altruism across societies.* In the third-party punishment game, the average acceptance threshold (the minimum offer not punished by the third party) provides a measure of the level of punishment in that society. This measure of punishment also predicts the level of altruism as measured by dictator generosity (the size of offers in the ordinary dictator game) across societies.

Taken together, these experiments indicate how institutions (e.g., punishment) and culture can strongly affect individual decision making.

Research on concepts of number and geometry show an interesting pattern of cultural variability and invariance in human concepts. Studies of several remote human groups provide evidence that verbal counting systems are not universal: there are languages that lack words for any exact numerical values beyond "two" or "three," and native speakers of those languages do not spontaneously represent exact numbers beyond that limit (Pica et al. 2004; Gordon 2004). Nevertheless, these speakers' abilities to represent approximate numerosities, and to perform approximate addition and subtraction on these representations, are indistinguishable from those of adults and children whose language and culture includes a full verbal number system and who have learned symbolic arithmetic. This suggests that the core representations of number that emerge in infants are universal across human cultures and provides a common foundation for culturally variable, higher level skills. A similar set of findings has been obtained with regard to geometric representations (Dehaene et al. 2006). If explicit calculations play a fundamental role in decision making, then the effect of culture on numerical concepts may be expected to have a fundamental impact on decision making. If decision making depends, however, on implicit calculations through the culturally invariant systems of number representation, as some evidence suggests (e.g., Chen et al. 2006), then patterns of decision making should show cultural invariance as well.

Intercultural comparisons that attempt to correlate differences in numerical concepts with decisions in games may be a fruitful way to test these ideas. More generally, experimental integration of cross-cultural, developmental, and cross-species comparisons may be extremely fruitful in specifying the elements of decision making.

Rate of Change

An important challenge to human decision making that directly relates to culture is the timescale over which biological and cultural evolution proceeds. Biological evolution gave rise to the neural mechanisms underlying the power of human cognition. The development of these mechanisms seems to have occurred relatively rapidly during biological evolution. However, the technological and social innovations to which these structures have given rise (i.e., the process of cultural evolution) have developed even faster and may outpace the process of biological evolution (Cohen 2005). Changes in the physical and social environment produced by cultural evolution may now be creating circumstances for which the older mechanisms in our brains are not fully adapted. This, in turn, may account for many of the idiosyncrasies and/or inconsistencies we observe in human behavior. Many modern circumstances may confront the human brain with decisions that can be made both by older mechanisms that were not adapted to the present circumstance as well as the more recently evolved mechanisms that support higher cognitive faculties which, in turn, could lead to conflict and inconsistency in behavior: competing responses to the unlimited availability of high caloric foods from our appetitive systems and from those that can take account of longer-term goals and consequences (Burnham and Phelan 2000).

In view of these rapid changes, there is a positive message in that humans have specifically evolved to respond to fast changes. It is likely that human populations have been subjected to rapid, high amplitude environmental changes for at least the last 800,000 years. Data from the Greenland ice caps provides a record of world temperatures over the last 120,000 years at a minimum resolution of ten years. Unlike the current interglacial climate period, these records indicate that glacial periods were characterized by fluctuations of world temperature from glacial cold to almost interglacial warmth, often over a period of a few hundred years. Lower resolution records from deep-sea cores suggest that human populations have been subjected to this varying climate regime for at least the last 800,000 years. It is plausible that the evolution of the capacity for social learning is an adaptation to such climate fluctuations, though of course climatic variation is but one element of the environment shaping cognitive adaptations. Unlike any other mammal, humans have the capacity to learn complicated skills easily by observing others. This is likely to be a refinement of the ability of primates to learn by observations, the neural basis of which may rely on mirror neurons that are found both in humans and monkeys. The much more efficient capacity of humans to learn skills through observation, however, allows for the rapid, cumulative evolution of locally adaptive skills—a trick that allowed human foragers to occupy almost every terrestrial habitat. Human intelligence alone has not equipped us to learn to live in such a diverse range of habitats, rather social learning has allowed human populations to accumulate gradually, but rapidly (by the standards of genetic evolution), highly habitat-specific technologies and knowledge. It is impossible, for example, for a single individual to find out how to make a kayak in the Arctic or a blow gun in the African desert without the benefit of the long line of cultural exchange between teachers and students. It has been argued that this specific adaptation may be both the source of our fast cultural evolution and the preparation to address the resulting challenges (see Boyd and Richerson, this volume).

Conclusions

Our uniquely human capacities have created a dense web of social interactions and a complex, rapidly changing cultural environment. Human culture differs in space and time, as do differences in explicit mathematical skills and institutions (e.g., morals, punishment and mandatory social securities). These differences have been shown to have dramatic effects on decision making. Examining the pathways in which culture interacts with reflective or reflexive processes within individuals and developing a better understanding for the dynamics of the complex interactions occurring in social networks remain important topics for further investigation. A better understanding of decision making at all levels will have important implications for our legal system, as it will inform our understanding of concepts such as responsibility and negligence.

Acknowledgments

We wish to thank Stan Dehaene for his precious input and Julia Lupp for her invaluable help in generating this report. Christian Keysers was supported by a Marie Curie Excellence Grant of the European Commission and a VIDI grant of the dutch N.W.O.

References

Allport, D. A. 1982. Attention and performance. In: New Directions in Cognitive Psychology, ed. G. I. Claxton. London: Routledge and Keagan-Paul.

Aziz-Zadeh, L., L. Cattaneo, M. Rochat, and G. Rizzolatti. 2005. Covert speech arrest induced by rTMS over both motor and nonmotor left hemisphere frontal sites. *J. Cogn. Neurosci.* **17**:928–938.

Baddeley, A. 1981. The concept of working memory: A view of its current state and probable future development. *Cognition* **10(1-3)**:17–23.

Baker, C. I., J. Liu, L. L. Wald, K. K. Kwong, T. Benner et al. 2007. Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proc. Natl. Acad. Sci.* **104(21)**:9087–9092.

Bloom, P. 2004. Descartes' Baby: How the Science of Child Development Explains What Makes Us Human. New York: Basic Books.

Bolger, D. J., C. A. Perfetti, and W. Schneider. 2005. Cross-cultural effect on the brain revisited: Universal structures plus writing system variation. *Hum. Brain Map.* **25(1)**:92–104.

Botvinick, M., and D. C. Plaut. 2004. Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychol Rev.* **111(2)**:395–429.

Brandimonte, M. A., J. W. Schooler, and P. Gabbino. 1997. Attenuating verbal overshadowing through visual retrieval cues. *J. Exp. Psychol.: Learn. Mem. Cogn.* **23**:915–931.

Braver, T. S., and J. D. Cohen. 2000. On the control of control: The role of dopamine in regulating prefrontal function and working memory. In: Control of Cognitive Processes, ed. S. Monsell and J. Driver, pp. 713–737. A Bradford Book. Cambridge MA: MIT Press.

Braver, T. S., J. R. Reynolds, and D. I. Donaldson. 2003. Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron* **39(4)**:713–726.

Burnham, T., and J. Phelan. 2000. Mean Genes: From Sex to Money to Food, Taming Our Primal Instincts. Cambridge, MA: Perseus Publ.

Camerer, C. F. 2003. Behavioral Game Theory: Experiments on Strategic Interaction: Princeton, NJ: Princeton Univ. Press.

Chen, M. K., V. Lakshminaryanan, and L. R. Santos. 2006. The evolution of our preferences: Evidence from capuchin monkey trading behavior. *J. Pol. Econ.* **114(3)**:517–537.

Cheney, D. L., and R. M. Seyfarth. 1990. How Monkeys See the World. Chicago: Univ. of Chicago Press.

Cohen, J. D. 2005. The vulcanization of the human brain: A neural perspective on interactions between cognition and emotion. *J. Econ. Persp.* **19(4)**:3–24.

Cohen, J. D., T. S. Braver, and R. C. O'Reilly. 1996. A computational approach to prefrontal cortex, cognitive control, and schizophrenia: Recent developments and current challenges. *Phil. Trans. Roy. Soc. Lond. B* **351(1346)**:515–1527.

Cohen, J. D., and R. C. O'Reilly. 1996. A preliminary theory of the interactions between prefrontal cortex and hippocampus that contribute to planning and prospective memory. In: Prospective Memory: Theory and Applications, ed. M. Brandimonte, G. O. Einstein, and M. A. McDaniel, pp. 267–295. Hillsdale, NJ: Erlbaum.

Daw, M. I., M. C. Ashby, and J. T. R. Isaac. 2007. Coordinated developmental recruitment of latent fast spiking interneurons in layer IV barrel cortex. *Nature Neurosci.* **10(4)**:453–461.

Dehaene, S. 1997. The Number Sense. New York: Oxford Univ. Press.

Dehaene, S. 2007. Reading in the Brain. New York: Penguin, in press.

Dehaene, S., V. Izard, P. Pica, and E. S. Spelke. 2006. Core knowledge of geometry in an Amazonian indigene group. *Science* **311**:381–384.

Donald, M. 1991. Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition. Cambridge, MA: Harvard Univ. Press.

Donald, M. 1993. Précis of "Origins of the Modern Mind" with multiple reviews and author's response. *Behav. Brain Sci.* **16**:737–791.

Donald, M. 1995. The neurobiology of human consciousness: An evolutionary approach. *Neuropsychologia* **33**:1087–1102.

Donald, M. 1998. Mimesis and the executive suite: Missing links in language evolution. In: Approaches to the Evolution of Language: Social and Cognitive Bases, ed. J. R. Hurford, M. Studdert-Kennedy, and C. Knight, pp. 44–67. Cambridge: Cambridge Univ. Press.

Donald, M. 1999. Preconditions for the evolution of protolanguages. In: The Descent of Mind, ed. M. C. Corballis and I. Lea, pp. 120–136. Oxford: Oxford Univ. Press.

Donald, M. 2001. A Mind So Rare: The Evolution of Human Consciousness. New York: W. W. Norton.

Donald, M. 2005. Imitation and mimesis. In: Perspectives on Imitation: From Neuroscience to Social Science, ed. S. Hurley and N. Chater. Cambridge, MA: MIT Press.

Elman, J. L. 1990. Finding structure in time. *Cogn. Sci.* **14**:179–211.

Farroni, T., G. Csibra, F. Simion, and M. H. Johnson. 2002. Eye contact detection in humans from birth. *Proc. Natl. Acad. Sci.* **99(14)**:9602–9605.

Feigenson, L., S. Dehaene, and E. S. Spelke. 2004. Core systems of number. *Trends Cogn. Sci.* **8**:307–314.

Fiebach, C. J., and R. I. Schubotz. 2006. Dynamic anticipatory processing of hierarchical sequential events: A common role for Broca's area and ventral premotor cortex across domains? *Cortex* **42**:499–502.

Frank, M. J., B. Loughry, and R. C. O'Reilly. 2001. Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cogn. Aff. Behav. Neurosci.* **1(2)**:137–160.

Gordon, P. 2004. Numerical cognition without words: Evidence from Amazonia. *Science* **306**:496–499.

Greene, J., and J. Haidt. 2002. How (and where) does moral judgment work? *Trends Cogn. Sci.* **6(12)**:517–523.

Greenwald, A. G., D. E. McGhee, and J. L. K. Schwartz. 1998. Measuring individual differences in implicit cognition: The implicit association test. *J. Pers. Soc. Psychol.* **74**:1464–1480.

Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychol. Rev.* **108**:814–834.

Hare, B., J. Call, and M. Tomasello. 2001. Do chimpanzees know what conspecifics know? *Anim. Behav.* **61**:139–151.

Hauser, M. D., H. Pearson, and D. Seelig. 2002. Ontogeny of tool use in cotton-top tamarins (*Saguinus oedipus*): Innate recognition of functionally relevant features. *Anim. Behav.* **64**:299–311.

Hauser, M. D., F. Tsao, P. Garcia, and E. S. Spelke. 2003. Evolutionary foundations of number: Spontaneous representation of numerical magnitudes by cotton-top tamarins. *Proc. Roy. Soc. Lond. B* **270**:1441–1446.

Henrich, J. 2000. Does culture matter in economic behavior? Ultimatum game bargaining among the Machiguenga. *Am. Econ. Rev.* **90(4)**:973–979.

Henrich, J., S. Bowles, C. Camerer, E. Fehr, and H. Gintis. 2004. The Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-scale Societies. New York: Oxford Univ. Press.

Henrich, J., R. Boyd, S. Bowles, H. Gintis, E. Fehr et al. 2005. "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behav. Brain Sci* **28:**795–855.

Henrich, J., R. McElreath, A. Barr, A. Ensimger, J. Barrett et al. 2006. Costly punishment across human societies. *Science* **312**:1767–1770.

Hochreiter, S., and J. Schmidhuber. 1997. Long short-term memory. *Neural Comp.* **9(8)**:1735–1780.

Hunt, G. R., M. C. Corballis, and R. D. Gray. 2006. Design complexity and strength of laterality are correlated in New Caledonian crows' pandanus tool manufacture. *Proc. Roy. Soc. Lond. B* **273**:1127–1133.

Jabbi, M., M. Swart, and C. Keysers. 2007. Empathy for positive and negative emotions in the gustatory cortex. *Neuroimage* **34**:1744–1753.

Keysers, C., and V. Gazzola. 2006. Towards a unifying neural theory of social cognition. *Prog Brain Res.* **156**:379–401.

Keysers, C., and D. I. Perrett. 2004. Demystifying social cognition: A Hebbian perspective. *Trends Cogn Sci.* **8**:501–507.

Kurzban, R., and C. A. Aktipis. 2007. Modularity and the social mind: Why social psychologists should be less self-ish. *Pers. Soc. Psychol. Rev.* **11**:131–149.

Landau, B., H. Gleitman, and E. S. Spelke. 1981. Spatial knowledge and geometrical representation in a child blind from birth. *Science* **213**:1275–1278.

Lemer, C., S. Dehaene, E. S. Spelke, and L. Cohen. 2003. Approximate quantities and exact number words: Dissociable systems. *Neuropsychologia* **41**:1942–1958.

Li, G., R. T. F. Cheung, J. H. Gao, T. M. Lee, L. Tan et al. 2006. Cognitive processing in Chinese literate and illiterate subjects: An fMRI study. *Hum. Brain Map.* **27**:144–152.

Matsuzawa, T. 1985. Use of numbers by a chimpanzee. *Nature* **315**:57–59.

McCandliss, B. D., L. Cohen, and S. Dehaene. 2003. The visual word form area: Expertise for reading in the fusiform gyrus. *Trends Cogn. Sci.* **7**:293–299.

McClelland, J. L., B. L. McNaughton, and R. C. O'Reilly. 1995. Why there are complementary learnings systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* **102(3)**:419–457.

McCrink, K., and K. Wynn. 2004. Large-number addition and subtraction in infants. *Psychol. Sci.* **15**:776–781.

Melcher, J., and J. W. Schooler. 1996. The misremembrance of wines past: Verbal and perceptual expertise differentially mediate verbal overshadowing of taste. *J. Mem. Lang.* **35**:231–245.

Melcher, J. M., and J. W. Schooler. 2004. Perceptual and conceptual expertise mediate the verbal overshadowing effect in a training paradigm. *Mem. Cogn.* **32(4)**:618–631.

Meltzoff, A. N., and W. Prinz. 2002. The Imitative Mind: Development, Evolution, and Brain Bases. Cambridge: Cambridge Univ. Press.

Mithen, S. J. 1996. The Prehistory of the Mind: The Cognitive Origins of Art, Religion and Science. Cambridge: Cambridge Univ. Press.

Montague, P. R., P. Dayan, and T. J. Sejnowski. 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16**:1936–1947.

Nakamura, K., S. Dehaene, A. Jobert, D. Le Bihan, and S. Kouider. 2005. Subliminal convergence of Kanji and Kana words: Further evidence for functional parcellation of the posterior temporal cortex in visual word perception. *J. Cogn. Neurosci.* **17(6)**:954–968.

Navon, D., and D. Gopher. 1979. On the economy of the human processing system. *Psychol. Rev.* **86**:214–255.

Naya, Y., K. Sakai, and Y. Miyashita. 1996. Activity of primate inferotemporal neurons related to a sought target in pair-association task. *Proc. Natl. Acad. Sci.* **93**:2664–2669.

Nelson, K. 1996. Language in Cognitive Development: Emergence of the Mediated Mind. New York: Cambridge Univ. Press.

Newcombe, N., J. Huttenlocher, and A. Learmonth. 2000. Infants' coding of location in continuous space. *Infant Behav. Dev.* **22(4)**:483–510.

Paulesu, E., J. F. Demonet, F. Fazio, E. McCrory, V. Chanoine et al. 2001. Dyslexia: Cultural diversity and biological unity. *Science* **291(5511)**:2165–2167.

Pepperberg, I. M. 1987. Evidence for conceptual quantitative abilities in the African Grey parrot: Labeling of cardinal sets. *Ethology* **75**:37.

Perfect, T. J., L. J. Hunt, and C. M. Harris. 2002. Verbal overshadowing in voice recognition. *Appl. Cogn. Psychol.* **16**:973–980.

Petersson, K. M., A. Reis, S. Askelof, A. Castro-Caldas, and M. Ingvar. 2000. Language processing modulated by literacy: A network analysis of verbal repetition in literate and illiterate subjects. *J. Cogn. Neurosci.* **12**:364–382.

Pica, P., C. Lemer, V. Izard, and S. Dehaene. 2004. Exact and approximate arithmetic in an Amazonian indigene group. *Science* **306(5695)**:499–503.

Posner, M. I., and C. R. Snyder. 1975. Attention and cognitive control. In: Information Processing and Cognition, ed. R. L. Solso. Hillsdale, NJ: Erlbaum.

Ratcliff, R., and G. McKoon. 1988. A retrieval theory of priming in memory. *Psychol. Rev.* **95**:385–408.

Ratcliff, R., and G. McKoon. 1996. Biases in implicit memory tasks. *J. Exp. Psychol.: Gen.* **125**:403–421.

Ratcliff, R., and G. McKoon. 1997. A counter model for implicit priming in perceptual word identification. *Psychol. Rev.* **104**:319–343.

Rizzolatti, G., and G. Luppino. 2001. The cortical motor system. *Neuron* **31**:889–901.

Sanfey, A. G., J. K. Rilling, J. A. Aronson, L. E. Nystrom, and J. D. Cohen. 2003. The neural basis of economic decision making in the ultimatum game. *Science* **300(5626)**:1755–1757.

Scholl, B. J. 2001. Objects and attention: The state of the art. *Cognition* **80**:1–46.

Schooler, J. W. and T. Y. Engstler-Schooler. 1990. Verbal overshadowing of visual memories: Some things are better left unsaid. *Cogn. Psychol.* **17**:36–71.

Schooler, J. W., S. Ohlsson, and K. Brooks. 1993. Thoughts beyond words: When language overshadows insight. *J. Exp. Psychol.* **122**:166–183.

Shah, P., and A. Miyake. 1996. The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *J. Exp. Psychol.: Gen.* **125(1)**:4–27.

Shallice, T. 1988. From Neuropsychology to Mental Structure. New York: Cambridge Univ. Press.

Shiffrin, R. M., and W. Schneider. 1977. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychol. Rev.* **84**:127–190.

Spelke, E. S. 2003a. Core knowledge. In: Attention and Performance, vol. 20, Functional Neuroimaging of Visual Cognition, ed. N. Kanwisher and J. Duncan. Oxford: Oxford Univ. Press.

Spelke, E. S. 2003b. What makes us smart? Core knowledge and natural language. In: Language in Mind, ed. D. Gentner and S. Goldin-Meadow. Cambridge, MA: MIT Press.

Squire, L. R. 1987. Memory and Brain. New York: Oxford Univ. Press.

Stanescu, R., P. Pinel, P.-F. van de Moortele, D. LeBihan, L. Cohen et al. 2000. Cerebral bases of calculation processes: Impact of number size on the cerebral circuits for exact and approximative calculation. *Brain*: **123**:2240–2255.

Stewart, L., R. Henson, K. Kempe, V. Walsh, R. Turner et al. 2003. Brain changes after learning to read and play music. *NeuroImage* **20**:71–83.

Tomasello, M. 1999. The Cultural Origins of Human Cognition. Cambridge, MA: Harvard Univ. Press.

Valenza, E., I. Leo, L. Gava, and F. Simion. 2006. Perceptual completion in newborn human infants. *Child Dev.* **77(6)**:1815–1821.
Vinckier, F., S. Dehaene, A. Jobert, J. P. Dubus, M. Sigman et al. 2007. Hierarchical coding of letter strings in the ventral stream: Dissecting the inner organization of the visual word-form system. *Neuron* **55(1)**:143–156.
Vohs, K. D., and J. W. Schooler. 2007. The Value of Believing in Free Will: Encouraging a Belief in Determinism Increases Cheating." *Psychol. Sci.*, in press.
Wicker, B., C. Keysers, J. Plailly, J. P. Royer, V. Gallese et al. 2003. Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron* **40(3)**:655–664.
Wilson, T. D., and J. W. Schooler. 1991. Thinking too much: Introspection can reduce the quality of preferences and decisions. *J. Pers. Soc. Psychol.* **60**:181–192.
Wynn, K. 1992. Children's acquisition of the number words and the counting system. *Cogn. Psychol.* **24**:220–251.
Xu, F., and E. S. Spelke. 2000. Large number discrimination in 6-month-old infants. *Cognition* **74**:B1–B11.

12

How Evolution Outwits Bounded Rationality

The Efficient Interaction of Automatic and Deliberate Processes in Decision Making and Implications for Institutions

Andreas Glöckner

Max Planck Institute for Research on Collective Goods, 53113 Bonn, Germany

Abstract

Classic behavioral decision research has intensively explored deliberate processes in decision making. Accordingly, individuals are viewed as bounded rational actors who, because of cognitive limitations, use simple heuristics that are successful in certain environments. In this chapter, it is postulated that human cognitive capacity is less severely limited than has previously been assumed. When automatic processes are considered, one finds that cognitive capacity is not a binding constraint for many decision problems. The general parallel constraint satisfaction (PCS) approach is outlined, which aims at describing these automatic processes, and evidence supporting this approach is summarized. It is further argued, that in order to describe decision making comprehensively, models must account for the interaction between automatic and deliberate processes. The PCS rule is delineated which specifies this interaction. The model shifts the bounds of rationality considerably and has further evolutionary advantages. Implications for the efficient design of institutions are outlined. Finally, the German legal system is reviewed in terms of its ability to support efficient decision making by implementing many of the prescriptions derived from the PCS rule without explicit knowledge about the underlying processes.

Introduction

One of the most intriguing psychological phenomena is the human ability to make decisions in a complex and uncertain world. Decision experts, such as