

Andreas Roepstorff · Chris Frith

## What's at the top in the top-down control of action? Script-sharing and 'top-top' control of action in cognitive experiments

Accepted: 11 August 2003 / Published online: 5 February 2004  
© Springer-Verlag 2004

**Abstract** The distinction between bottom-up and top-down control of action has been central in cognitive psychology, and, subsequently, in functional neuroimaging. While the model has proven successful in describing central mechanisms in cognitive experiments, it has serious shortcomings in explaining how top-down control is established. In particular, questions as to what is at the top in top-down control lead us to a controlling homunculus located in a mythical brain region with outputs and no inputs. Based on a discussion of recent brain imaging experiments, we argue for the need to factor the interaction between the experimenter and the experimental participant into a realistic understanding of top-down control. We suggest these interactions involve a 'sharing of scripts' for perception and action that may be described as 'top-top processes.' We thereby expand the understanding of the homunculus to include elements of social cognition. This conceptual reconfiguration may grant some sort of asylum for a—not very omnipotent—homunculus.

**Abbreviations** *WCST* Wisconsin card sorting task · *DLPFC* Dorsolateral prefrontal cortex · *RPS* Rock-paper-scissors · *CRT* Choice reaction time task

### Metaphors of up and down in the brain

The distinction between 'top-down' and 'bottom-up' processes is one of the more important conceptual models in cognitive psychology. This organization of neuronal and cognitive processes along a vertical axis appears very straightforward, indeed almost intuitive, but on second thoughts, things may get more complicated. Along which kind of dimension does this movement occur?

On the one hand, there appears to be a purely anatomical distinction between 'lower' and 'higher' parts of the brain, where 'the top' of this axis is to be understood, somewhat loosely, as the anterior frontal regions such as the prefrontal cortex, and the bottom are the more posterior and 'deeper' structures terminating at the brain stem. In these terms, the model also carries evolutionary connotations, as the axis distinguishes the 'lower' reptile brain from the higher, more advanced mammalian—and ultimately human—brain.

The model also refers to a relationship between the organism and the environment. In this sense, 'bottom' is that which comes to the organism 'from the outside.' This is usually taken to be the sensory inputs, while 'the top' is all that understanding and knowledge that the organism has on its own (Frith & Dolan, 1997). Finally, 'top' and 'down' also evoke semantic resonances of power and control, as in the colloquial distinctions between 'the top executive' and 'the man on the floor' (Lakoff & Johnson, 1980).

Perhaps the beauty, and indeed also effectiveness, of the top-bottom model for describing both brain organization and cognitive function is precisely that it brings together these very different levels of organization: An anatomical organization, an evolutionary perspective and notions of control and governance. However, as with most other metaphorical relationships that serve to structure how thoughts are expressed and how they are set to work with each other (Lakoff & Johnson, 1980), there comes a point where the reality described seems to

A. Roepstorff  
Center for Functionally Integrative Neuroscience,  
Aarhus University, Building 30, 44 Nørrebrogade,  
8000 Aarhus, Denmark

C. Frith (✉)  
Wellcome Department of Imaging Neuroscience,  
University College London, 12 Queen Square,  
London, WC1N 3BG, UK  
E-mail: cfrith@fil.ion.ucl.ac.uk  
Tel.: +44-20-78337472  
Fax: +44-20-78131420

resist the apparent simplicity of the metaphorical model. We believe that this becomes apparent once one considers how the notions become implemented in the design and interpretation of cognitive experiments.

### Where is the top in top-down control?

The distinction between ‘bottom-up’ and ‘top-down’ control of action is nicely illustrated in response selection tasks. In a choice reaction time task (CRT) where participants have to move a finger as soon as that finger is touched the control of response selection is bottom-up since the imperative signal (in the environment) indicates which response should be selected in each trial. In contrast to other CRTs in which the relationship between stimulus and response has to be learned, the responses in this tactile version are automatic and reaction time does not increase with the number of possible choices from 2 to 8 (Leonard, 1959). In a willed action task the control of response selection is top-down since, in each trial, it is the participant who has to decide which response to select. The imperative signal merely indicates when the response should be made. The flow of information that controls response selection in these two cases is illustrated in Fig. 1.

By using functional brain imaging it has been possible to assign brain regions to some of the boxes in these diagrams. In particular, top-down response selection is consistently associated with activity in the prefrontal cortex (Deiber et al., 1991; Frith, Friston, Liddle, & Frackowiak, 1991).

There are several problems with this diagram of top-down control. First of all, if we take the box-and-arrow diagram seriously as a model of what happens at the neural level it tells us that there should be a region in the brain (the top) in which there are only outputs and no inputs. There is no evidence that such a region exists (S. Zeki, personal communication). Secondly, if we think about what is at the top in cognitive terms we are

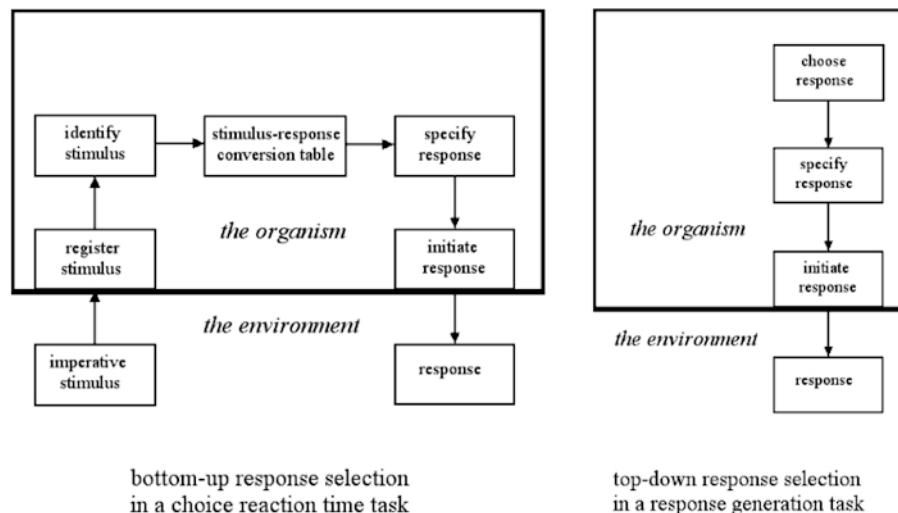
confronted with the problem of what or who makes the response selection. Presumably it is the much-maligned homunculus. Hence, the logical consequence of the representation in Fig. 1 is that the homunculus must be located in a place in the brain that has only output and no inputs. This place is a bit like the country east of the Sun and west of the Moon. It is therefore no wonder that the dominant trend within cognitive science has been to declare the homunculus as real as the hobbit or other fairy tale creatures.

We believe, however, that the problem with the homunculus may be conceptual rather than ontological, i.e., the blame may be on idealized models of cognitive function such those depicted in Fig. 1 rather than on that unreal homunculus arising from it. In the following we will examine in some detail recent brain imaging experiments on top-down control of action. We shall argue that models such as that depicted in Fig. 1 will not suffice to account for how the observed pattern of action is installed in the first place. We will suggest that a notion of ‘shared scripts for action’ may remedy some of these shortcomings. This will allow us to create a space of asylum for a homunculus that is, at the same time, more stupid, and somewhat less omnipotent and consistent than the one implicated in Fig. 1.

### Cross-species neural correlates of action

In a recent article in *Science*, Nakahara, Hayashi, Konishi, and Miyashita (2002) described a remarkable experiment. Not only had the authors succeeded in training two macaque monkeys to perform a simplified version of the Wisconsin card sorting task (WCST), they had also managed to familiarize the animals with all the unpleasantness of a running MR scanner—the noise, the fixation of the head, the limited field of view etc.—to such a degree that the monkeys could perform the task while undergoing a fMRI examination.

**Fig. 1** Bottom-up and top-down processes in cognitive tasks



As is probably well known, the experimental participant in a WCST is presented with cards that display symbols in specific shapes, colors, numbers, etc., such as three green circles, or four yellow triangles. The task of the participant is to sort the cards into different piles without knowing the criteria for a correct sorting. They are given feed-back about the correctness of their sort after each card has been placed, and once they have discovered the sorting rule, e.g., that the cards should be sorted by color, the sorting dimension is changed by the experimenter, and the participant then has to discover the new rule, e.g., that the cards are to be sorted by shape. Throughout decades of research (Milner, 1963; Monchi, Petrides, Petre, Worsley, & Dagher, 2001; Stuss et al., 2000), the WCST has been established as a standard cognitive test that targets participants' abilities to switch between cognitive sets, and a poor performance in the test can be a key neuropsychological indicator of a putative prefrontal lesion. The control of action needed in this task is clearly top-down.

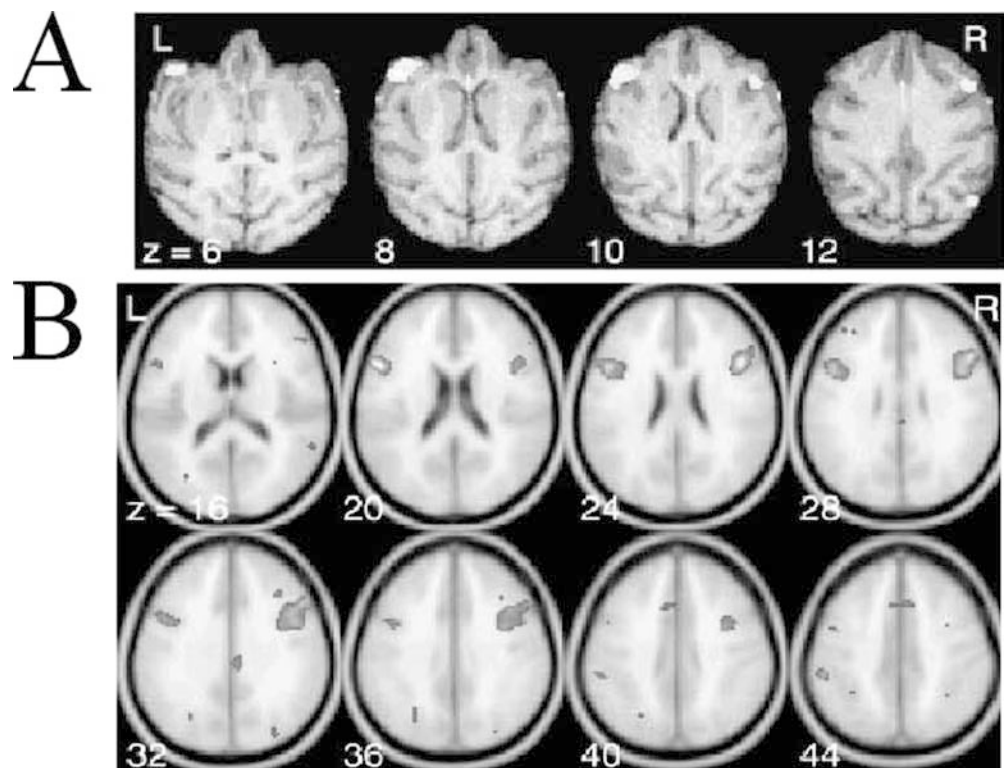
Nakahara et al. trained monkeys to perform a computerized version of the WCST where a target card was displayed on the computer screen, followed by three cards that differed on two dimensions (color and shape of symbols). The monkeys were then supposed to select one of the three cards that reflected the current sorting criteria (color or shape). Feedback was provided by a visual display on the screen and a liquid reward for the correct choice.

The main effect of the set-shifting component of the WCST was found in the ventrolateral prefrontal cortex bilaterally, (see Fig. 2A) at the ventral end of the inferior

ramus of the arcuate sulcus. Subsequently, the same experiment was conducted on 10 human volunteers—the only difference being that they were not given a liquid reward following a correct answer. In these participants, the main activation was found in the posterior part of the bilateral inferior frontal sulcus (Brodmann's area 44/45, see Fig. 2B). This activation site has also been found in other studies on the WCST. Nakahara et al. argue that the main sites of activation in the two species, the Brodmann area 44/45 in humans and the Petrides and Pandya area 44/45 in macaque monkeys have a similar cytoarchitecture. The authors, hence, suggest that they may be considered both functional and anatomical homologues.

We think that the experiment demonstrates two highly interesting findings. The first, and perhaps most obvious, is the elegant demonstration of inter-species functional and anatomical homologues in higher cognitive function. In order to perform well in a WCST, the experimental participant—be that human or monkey—must be able to establish, maintain, and alter a particular cognitive set, which can be used as a template for acting in the world. For that they appear to draw on activity in anatomically, ontogenetically, and functionally similar brain regions in the prefrontal cortex. This seems to be a clear case of a 'top-down' control of action in two senses of the metaphor discussed above. It is an 'executive top' that establishes a model for acting in the world, the efficacy of which is continuously monitored, and it is an 'anatomical top', since the main areas of activations seem to take place in very 'top-like' frontal regions of the brain. In this respect, the brain scans reveal no apparent differences between the two species.

**Fig. 2** Shift related activations elicited by the Wisconsin card sorting task (WCST) in **A** macaque monkeys and **B** humans. **A** Statistical parametric map (SPM) overlaid on a transverse section of a normalized, structural MR image. The main activations are in the posterior part of the ventrolateral prefrontal cortex bilaterally, approximately at the Petrides and Pandya area 44/45. **B** Main activations are bilaterally in the posterior part of the ventrolateral prefrontal cortex. Reproduced from (Nakahara et al. 2002)

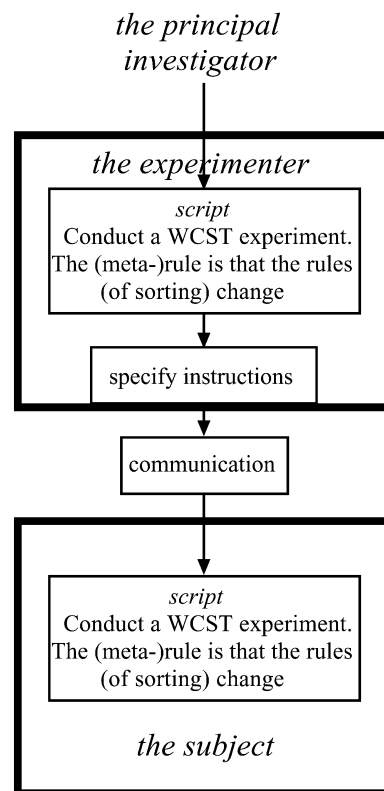


The details of the experiment reveal another equally interesting aspect of action control, but this one is not visible from the images depicting activated brains. Although both monkey and man managed to perform very well in the WCST at the time of testing, they had learned it in very different ways. Getting the macaque monkeys to do the trick proficiently in the MRI scanner involved about 1 year of operant training. Compared with this very arduous process, the human volunteers simply did the task according to the verbal instruction, with a relatively short (30–60 min) period of familiarization training before the MRI scan (Yashushi Miyashita, personal information). These differences are noteworthy in that they point to two very different ways of establishing similar—in terms of behavior and brain activation—patterns of control of action.

The human case may appear trivial. As a task, WCST is interesting because the important rule is a meta-rule, i.e., a rule about a rule (cf. Bateson's 1972 concept of meta-communication as 'communication about communication'). This meta-rule, which the participant has to understand in order to act correctly, says that 'the (meta-) rule of the game is that the rules (of sorting) keep changing.' In later versions of the task the participant gets this meta-rule from the verbal instructions. Once he has understood it, he knows how to interpret that stream of stimulus-response-feedback, which the actual experiment consists of, as instances of the underlying rule. This is clearly a case of top-down control, since this understanding of the situation allows the participant to act correctly—ideally even on the first experimental run.

Aspects of the 'top-down' metaphor break down at this point. As shown in Fig. 1, bottom-up processes are usually considered to be driven 'from the outside' by the sensory inputs, while top-down processes are driven 'from the inside' by mental processes. However, the 'verbal instructions' that enable the human volunteers to perform well in the task, fail to fit this scheme. The instructions are clearly coming 'from the outside' and are mediated via the senses, i.e., bottom-up, and yet their main purpose is to allow for the very rapid establishment of a consistent model for how the participants are to interpret and respond in the situation, i.e., top-down. We suggest that the solution to this conceptual problem is to factor the experimenter into the model of the control of action in the experiment.

The purpose of the instructions in this experiment and in general is that the experimental participant and the experimenter come to share a common understanding of the nature of the experiment and of the intended stimulus-response relationship (Roepstorff, 2001). We have suggested elsewhere (Jack & Roepstorff, 2002) that this process can be described as a 'sharing of scripts' between the experimenter and the participant (Fig. 3). In this model, the 'top' of the 'top-down' process is no longer a hypothetical place of outputs, without inputs (Fig. 1). Instead, the origin of the 'executive top' employed in the WCST is outside the brain of the participant, namely in the mind of the experimenter. We call



**Fig. 3** Communicating the WCST script through a 'top-top' exchange

this interaction between experimenter and experimental participant a 'top-top' exchange of scripts. Of course, the sharing of scripts depends upon instructions communicated through visual or acoustic signals. However, once this communication is successfully accomplished participant and experimenter have common, shared representations about the nature of the task (see Pickering & Garrod, in press for a similar argument about aligned representations in successful dialogue). These shared representations largely concern top-level aspects of control, i.e., the goals of the task rather than low-level aspects specifying precisely how movements should be made.

At this level of analysis, the contrast with the monkey, which must go through 1 year of operant training, is dramatic. We don't know much about what goes on in the monkey's mind during that period. Judging from the amount of training needed, it seems reasonable to believe that establishing an understanding of the experimental situation, which allows the monkey to act proficiently and to be rewarded at an 'acceptable' level—whatever that means for a monkey—must be a tough process. However, judging from the performance at the time of scanning, which was comparable to that of the human participants, it seems likely that the end-result is that the monkey comes to enact a script with strong affinities to the experimenter's script of the WCST: The (meta-) rule is that the rule (of sorting) constantly changes. The finding by Nakahara et al., that

there were anatomically similar brain activations in humans and monkeys, supports this interpretation. In this respect, the two experimental scenarios are similar. But whereas the human participant receives this script directly from the experimenter in a 'top-top' exchange, the monkey has to reconstruct this script solely via the concrete stimuli and rewards offered to it. It happens as the monkey, based on the previous understandings of the situation, reacts to the reward responses that the experimenter dispenses. We propose to call this set of interactions a top-down-bottom-up process of synchronizing scripts (Fig. 4).

In our use of the 'script' metaphor, we have been inspired by the world of theatre; by the relation between the performance of an actor enacting a particular role on a stage and the underlying script (Jack & Roepstorff, 2002). It seems likely to us that both monkeys and humans in the experiment by Nakahara et al. came to share aspects of the experimenter's script for the experimental situation. However, the details of the experimental setup suggest that there are also important differences in the scripts enacted by the two species. The human participant knows that he is taking part in an experiment and that this requires that he should expose himself to the somewhat unpleasant environment of a MR scanner without complaining and without moving (Roepstorff, 2002). This contrasts with the experimental procedure for the monkey where the head is fixed with an implanted head holder. The feed-back structure in the experiment is also very different between the species, the monkeys are provided with liquid rewards when they sort correctly, while the human participants do not need rewards, but only visual cues indicating whether they are right or wrong.

As with the differences between the 'top-top' and the top-down-bottom-up' exchange of scripts, this suggests that there is a much greater overlap between the script of the experimenter and that of the human participant than between the script of the experimenter and that of the monkey. The differences can be summarized by saying that in the human case, it is expected that experimenter and participant come to share an interpretive frame for the context of the experiment, for the interpretation of the stimuli, and for the proper plans for action. Based on the article by Nakahara et al., we have no way of estimating the actual degree of script sharing between the experimenter and the monkey. But we may hypothesize that the ease with which an experimental set up with humans can be established points to one of the most interesting aspects of human perception, cognition and control of action: The almost automatic sharing of contexts, interpretive frames, and schemes for action.

### The script in 'willed action' tasks

We have discussed the role of scripts in the 'top-down' control of experiments, demonstrating that what is at the 'top' of most experiments is an interaction between

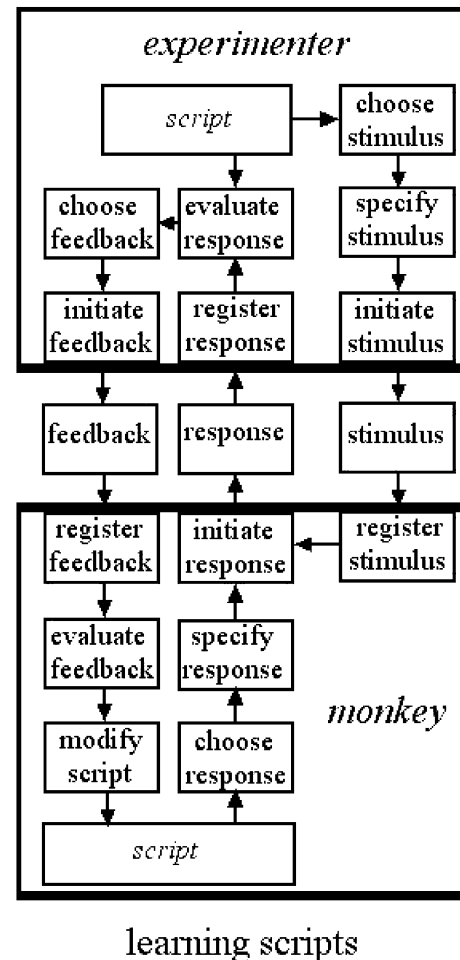


Fig. 4 Learning the WCST script via top-down-bottom-up feedback

the participant and the experimenter. Often it is the experimenter who is at the top in the sense that it is he who determines what the participant will do. We shall now return to a consideration of 'willed action' tasks and see whether the concept of scripts can also help us to understand top-down control in these tasks.

We believe that willed action tasks are extreme examples of top-down control. This is because the instructions are inadequate to determine what the participant should do. In effect the experimenter is saying to the participant, 'I am not going to tell you what to do. You must decide for yourself.' This is why we characterize such tasks as depending upon will. But by characterizing them in this way we seem unavoidably to introduce a homunculus into our cognitive model. At some point in the model an action has to be selected (see Fig. 1) without any input from the external environment. So something, a homunculus, has to be doing the choosing.

This idea of something that selects actions of its own will is fundamental to our concept of agency. If we observe that the actions of some creature are entirely pre-determined by forces or signals in the environment then

we do not consider that creature to be an agent. On the other hand, if we see behavior that is not simply responsive to changes in the environment we classify the creature as an agent. This applies to something as simple as a dot moving on a screen. If unexpected changes in speed or direction are observed we rate the dot as being more like an agent (Tremoulet & Feldman, 2000). A fundamental feature of our world is that it contains a class of things (agents) that can choose actions for themselves. It is also fundamental that each of us is one of those kinds of things. Beyond the consideration of motives in everyday life we do not concern ourselves too much with the precise mechanisms by which these agents can freely select actions.

So when we instruct someone to perform a willed action task we are effectively saying, 'Be an agent. Choose the responses for yourself.' But, of course, this instruction has to be unpacked and a script generated. To be like an agent we have to select our responses in such a way that an observer cannot easily predict what our next response will be. And the best way to do this is to choose the next response at random. In random number generation tasks, this instruction is made explicit, 'generate a series of random numbers as if you are taking them out of a hat.' The pattern of brain activity associated with explicit random number generation and willed action tasks is strikingly similar (Frith et al., 1991; Jahanshahi, Dirnberger, Fuller, & Frith, 2000). We believe it likely that, when participants are asked to select responses 'at will' they assume a script that requires them to respond randomly. This analysis seems to imply that the participant in a willed action experiment is not being a free agent, but is simulating being an agent by choosing responses at random.

But how do we generate a sequence of random responses? We could put numbers into a hat, shake it, pick out one number without looking and then repeat the process. Indeed, this is how the task was explained to participants in the random number generation experiment by Jahanshahi et al. (2000). Unfortunately, we can't do this mentally. We may have a mental image of the hat with the numbers in it, but it is hard to imagine picking out a piece of paper without knowing what number we are going to find on it and then being surprised by the result. So a different strategy has to be used based on a more complex script concerned with what we believe random sequences are like. Some of these beliefs are correct: Each number should occur equally often. Others are false (the gamblers fallacy): The same number won't occur twice in succession, numbers will not come in sequential order (3 after 4 or 4 after 3). These rules for randomness are used to constrain response selection in each trial. If I have chosen 1, 4, and 7 so far, in the next trial I must choose a different number that isn't 6 or 8. Also I had better break away from the upward trend and not choose 9. This leaves me with just 2, 3, or 5. Thus, even in a willed action task, most of the work in selecting responses is not made by the participant (i.e., the homunculus), but by the implicit script imposed by the

experimenter. Furthermore, once the acceptable responses have been narrowed down to a few possibilities, it doesn't really matter which of those remaining possibilities are chosen. Environmental triggers could be used to make this final choice, leaving nothing for the homunculus to do. If a choice must still be made, whatever does the choosing does not have to be smart.

We have argued elsewhere (Nathaniel-James & Frith, 2002) that the activity observed in the dorsolateral prefrontal cortex (DLPFC) during willed action and random number generation tasks is not associated with will (i.e., endogenous selection), but with the specification of which responses are acceptable and which are inappropriate given the script associated with the task (Fig. 5). When the cognitive load becomes too high, as in dual task situations, activity in the DLPFC is reduced and inappropriate responses are produced. Damage to the DLPFC has similar effects. Thus, this region of the frontal cortex that has long been associated with top-down control, would better be characterized as having a role in the implementation of scripts.

We have analyzed willed action tasks in some detail in order to demonstrate that even for these tasks, top-down control is better thought of in terms of control by the experimenter who in turn is controlled by scientific knowledge provided by his predecessors and peers. The difficulty with these tasks is that this control is very indirect. A lot of work has to be done by the participant

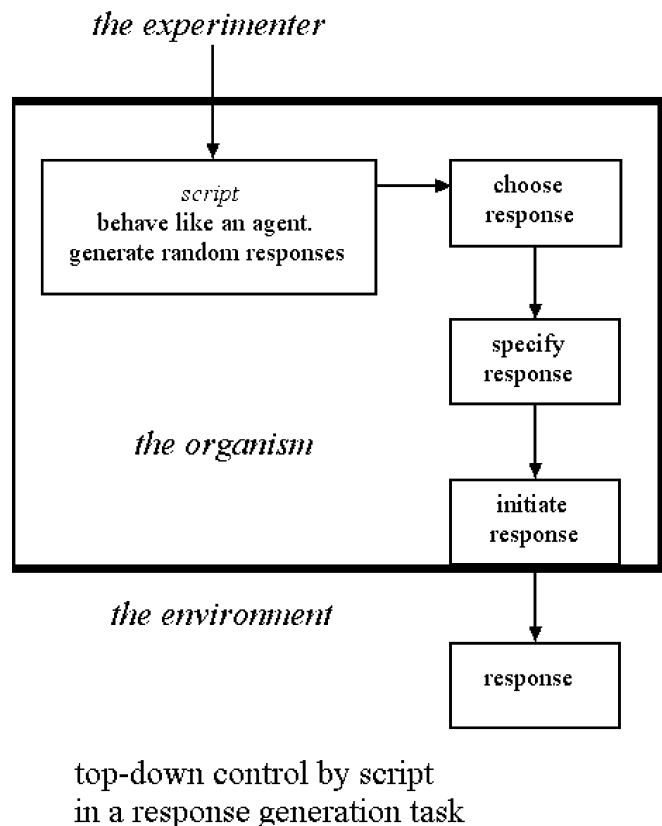


Fig. 5 Top-down control by script of a willed action task

to unpack the script implicit in the instruction to 'select the response for yourself.' The participant has to work out what the experimenter means by this instruction. The instruction, 'press the left button when you see the green light' is much more straightforward. It takes a long time to train monkeys to perform top-down tasks like the WCST precisely because monkeys cannot be told—and do not have the capacity to work out quickly—what the experimenter has in mind. Once the script has been understood through verbal instruction or trial and error learning, similar brain areas appear to be used by monkeys and humans to implement the script. But these are probably not the areas that are needed for understanding the script in the first place.

Performance of top-down tasks depends on an intimate interaction between experimenter and participant in which they successfully share a script for the performance of the task. Performance of the task is a joint enterprise, but it is only rarely described as such in the cognitive literature. Outside the laboratory this strange distinction between participant and experimenter no longer holds and the shared nature of our endeavors is much more obvious. Once the analytical attention is shifted from the actual performance of the participant during the experiment to the 'setting up' of the experiment, the inherently shared aspects of the cognitive experiment become very apparent (Roepstorff, 2002, 2003, *in press*). This approach marks a change of focus from the classical object of cognitive neuroscience, the enacting of the experimental script, to the standard object of analysis of ethnography and anthropology, the construction and framing of the performance (Jack & Roepstorff, 2002; Roepstorff, 2001). In a sense, it is—like the WCST discussed above—a shift from a primary level to a secondary or meta-level.

We have recently conducted a script-based brain imaging experiment into the ascription of agency that may serve to pinpoint this distinction.

---

### **Mentalizing and the experience of agency**

Gallagher, Jack, Roepstorff, and Frith (2002) report a PET experiment where the effective contrast was established by a computerized version of the game rock-paper-scissors (RPS). In the two main conditions, participants were led to believe that they played against either a human opponent or a computer that followed simple rules. In fact, apart from two short lead-in and lead-out sequences before and after the actual scanning window, they played against a random sequence.

Two types of data were collected from the experiment. Analysis of the cerebral blood-flow using PET showed a single area in the paracingulate cortex that was more strongly activated when participants believed they were playing against a person. No brain regions were differentially activated when participants believed they were playing against the computer. Retrospective reports, obtained by semi-structured interviews after the

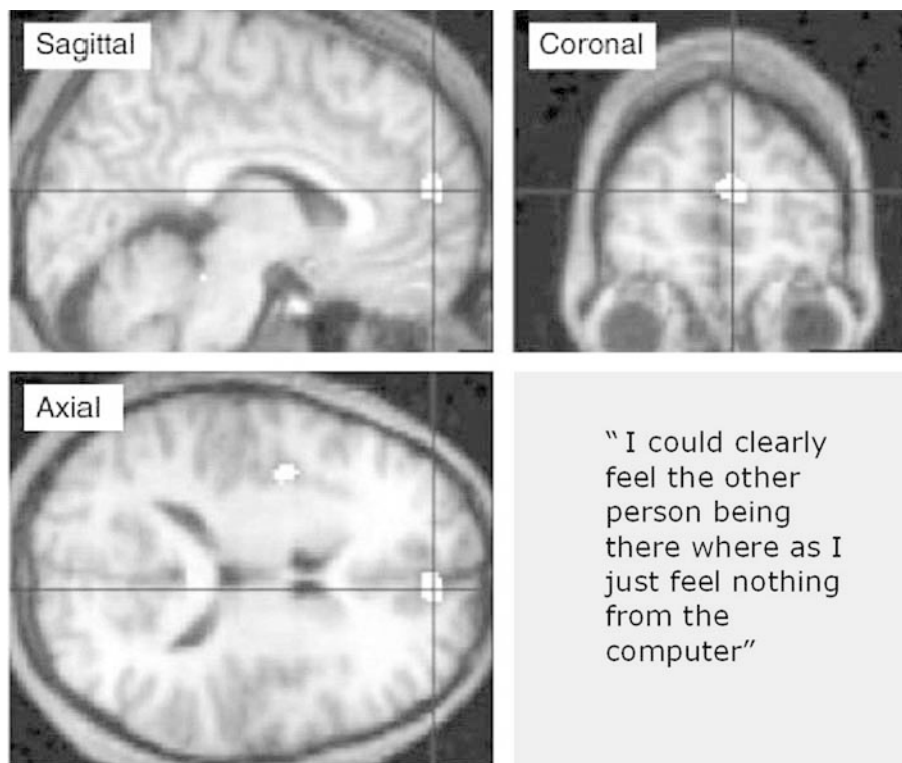
experiment, all described phenomenal differences between the two conditions. Briefly, in the conditions where participants believed they were playing against a person, they all described experiences of an intentional interaction with the opponent using words like guessing, double-bluffing, discovering individual strategies, and seeing the opponent's moves as 'something you can go along with.' In other words, all participants mentalized the imaginary opponent (Gallagher & Frith, 2003). In contrast, the computer was mainly described by participants as either random or rule-bound. We have argued (Jack & Roepstorff, 2002) that as the stimuli during the scanning window were identical in the two conditions, the significant contrast in the objective brain scanning measurements and the coherent differences in the phenomenal subjective reports were driven by differences in the script. These were mainly established via the instructions for the experiment and the prior knowledge the participants had of playing the game and of interacting with persons and computers.

We believe that the RPS experiment is a somewhat extreme case of top-top control of action in a cognitive experiment. As with the experiments on random number generation and the WCST, discussed above, the brain activation elicited is a result of the implementation of a particular script. In this experiment we take the observed brain activation to be related to that particular attitude or stance that the participant assumes in interpreting the stimuli and in determining his responses, rather than to the actual execution of the response. Adopting the script of playing against a person rather than the script of playing against a computer requires that the participant adopts an intentional stance towards his opponent. In other words, in this condition, not only must the participant share the script proposed by the experimenter, he must also think about the script adopted by his opponent.

In the post-experiment interviews, none of our participants expressed doubts that they had played a person or a computer. In a very interesting comment, one of our participants even reflected on the understanding of the experiment in the following way: "I clearly felt someone was there [when playing the person]. It is very difficult to see from the pattern [of the opponent's responses]. I [am] not sure that I would have had that sense, had I not known that I was not playing the computer." This suggests that once a particular stance has been taken, both the perceptual qualities and the actual button presses followed, in a sense without reflection (Fig. 6). There is nothing in the post-experiment interviews that speaks against this interpretation. Perhaps more anecdotally, it is also in accordance with the post-hoc reflections of the first experimental participant, who happens to be one of the authors of this article.

We consider that this experiment shows two interesting aspects of 'control of action.' Firstly, participants are in an almost trivial sense in control of their own actions during the experiment, nobody tells them which button to press when playing the computer or when

**Fig. 6** Typical phenomenal report and region of maximum activation in the paracingulate cortex elicited when participants 'take the intentional stance' in the RPS experiment. Redrawn from (Gallagher & Frith, 2003), images display group data mapped onto a template brain



playing the person. Yet their understanding of the situation, which motivates their actions, is largely governed by the frame of interpretation provided by the experimenter. If the participant distrusts this, the experiment is likely to fail. In other words, the experiment only works because the framing of it is not made explicit. Had the participants been involved in a meta-communication about the experiment prior to the examination: *We tell you that you play a person, but in fact, you play a random sequence*, we would have been surprised to find any activations.

Secondly, the experience of agency in the participant is closely linked to his understanding of agency in the opponent. Judging from the interviews, interacting with a person seems to invoke almost prototypical understanding of human agency as something that is neither rule-bound nor random. The actions of the other is something you may go along with and, during the interaction, you may develop an understanding of their particular characteristics. It is uncertain whether it is actually possible to learn to identify these psychological traits, or whether they really exist. Computer simulations of RPS interactions suggest that they may develop into Hamiltonian chaos situations, where the trajectory may be simple or complex, depending on the initial conditions (Sato, Akiyama, & Farmer, 2002). However, as argued by Sato et al. (2002), nothing indicates that actual players choose the 'rational' strategy, i.e., to play each move randomly. Theoretically, this strategy has the advantage that there is no strategy that can beat it, but the disadvantage that it cannot beat any strategy. Furthermore, as discussed above, it is probably impossible

to implement random choices in a human mind anyway. Instead, the default option seems to be to assume that the opponent, like oneself, is an agent with all that entails in terms of drives, motives, strategies, and aspirations. And it is this 'homunculus' that forms the template for action in the game.

---

#### **'Top-top' interactions, parallel processes, and consciousness**

The problem about standard representations of top-down functions such as depicted in Fig. 1 was that the ultimate source of control, 'the top,' becomes a free-floating independent homunculus. We believe that one important lesson to be drawn from the studies on 'willed action' and on 'mentalizing' discussed in this paper is that although the experimental participants in both situations report that they are in control of their actions, further analysis demonstrates that, in fact, both the actual patterns of behavior and the phenomenal experiences are influenced by top-top interactions on one hand, and by underlying, unattended processes on the other. This understanding does not go down very well with a notion of executive control by an omnipotent homunculus, but it may open up another interpretation.

Even when actions are determined by some shared script we still have the experience that there is something in us that makes endogenous choices and thus we continue to be an agent. Even when brain events run so fast that they cannot be attended to in real time, we still have the experience that they are ours. In contrast with the



homunculus assumed in representations such as Fig. 1, it seems that the actual capacities of the homunculus are, in fact, very limited. It is constantly left unaware of the actual links between perception and action, and it is constantly being overwhelmed by clever tricks imposed by other people. And yet it may be in relation to an entity like this that basic understandings of agency, of self, and of other are being linked. Instead of explaining the little man away, we propose to grant him some mental asylum. Perhaps we need this little virtual person not only to preserve the idea that we are in control of our own actions, but also to preserve the idea that we are stable selves, unchanged by all the influences thrust upon us by others.

We suggest that a useful conceptual space for a notion of the homunculus may be located at the nexus between those many parallel processes that the brain is constantly engaged in, and the input from other people, of top-top interactions. In this understanding, the role of a putative homunculus becomes one of a dual gate keeper: On one hand, between those many parallel processes and the attended few, on the other hand between one mind and another.

Wegner (2002) has claimed that the sense of being an autonomous self is an illusion. As we have argued above, the feeling of control and consistency may indeed seem illusionary from an outside perspective. However, from the inside perspective of the individual, it appears to be a very important anchor point both for action and perception. If we did not have the experience of this inner homunculus that is in control of our actions, our sense of self would dissolve into the culture that surrounds us.

A more important feature of this awareness is that it puts us in touch with other agents. It is this aspect that has been largely ignored in discussions of the neural correlates of consciousness. This is perhaps the most important message of the experiment reported by Nakahara et al. (2002). Being in touch with others, sharing frames of interpretations and models for action through 'top-top' interactions, is a prerequisite for setting up cognitive experiments fast and efficiently. Indeed, this ability appears to be piggybacking on a much more general human cognitive competence. We believe that humans differ from other animals most in the possession of a complex of abilities that allow for 'top-top' interactions. Within social anthropology, it is old hat that culture, understood as that which can be shared and exchanged, is an integral part of human nature. This implies that it may be difficult to separate 'social cognition' from 'nonsocial cognition' since so much of human attention and cognition is directed to, and informed by other people, and the cognitive experiment is but one interesting special case of this general condition (Roepstorff, 2001). It is, however, only recently that models of human cognition within a biological and evolutionary framework appear to converge on similar understandings (e.g., Deacon, 1997, Tomasello & Rakoczy 2003).

We are arguing that top-down control is not, after all, only control by the 'self,' but instead it is control med-

iated by other minds and more generally by culture. In the case of the experimenter the script is provided by the community of scientists who publish in his field (Fleck, 1979; Roepstorff, 2002, in press). While bottom-up control is exerted by the physical world, top-down control is exerted by the mental world. Typically actions under top-down control are determined by considerations like, 'what does he want me to do?' Or 'what is the appropriate thing to do in this situation?' An important feature of top-down control is that we are aware of selecting our actions, whereas bottom-up control can often be achieved without awareness. This is where the self comes in; because we are aware of selecting an action we feel that we are acting as an autonomous agent. Even in an experimental setting, we could have chosen to do something different such as playing tricks with the experimenter (Roepstorff, 2001).

Although we believe there are good reasons to maintain a 'weak' analytical notion of a homunculus, we do not know exactly in which part of the brain to grant it physical asylum. It has been argued that an 'attentional homunculus' should be conceived of as a parietal-frontal system rather than of one particular place (Nobre, 2001). We do not disagree with this. However, it is tempting to suggest that one putative anchor point for a homunculus, which serves as reference for the experience, ascription and detection of agency, is somewhere between the anterior part of the cingulate gyrus and the paracingulate gyrus. This area constantly turns up in a large variety of brain imaging studies related to attention to action, to the self, and to others (Frith, 2002; Gallagher & Frith, 2003; McCabe, Houser, Ryan, Smith, & Trouard, 2001; Vogeley et al., 2001). We have argued above that attention to action, to the self and to others can be thought of in terms of sharing scripts. One component of script sharing seems to be the ability to take the perspective of another person. Indeed, in a recent brain imaging study, which studied the contrast between simulating the perspective of another person and the self perspective, this region was activated (Ruby & Decety, 2001). There is furthermore some evidence that damage to this area can create a disorder known as 'environmental dependency syndrome' (Lhermitte, 1986). Actions of patients suffering from this condition appear to be determined by highly stereotyped and common scripts, which, almost bypassing 'free will', appear to be triggered by objects present in the immediate environment, e.g., seeing a bed in the neurologist's flat prompts the patient to take off his clothes and go to sleep.

These findings all appear in accordance with a model of action control where neuronal activity in the anterior cingulate/paracingulate may serve as one anchor point for the ascription of agency and script both in the self and in others. For those who fancy evolutionary 'just so' stories it may, furthermore, be relevant that certain parts of the anterior cingulate appear to have undergone major morphological changes in the recent evolution of pongids and homids, including the development of an unusual type of large spindle shaped projection neurons

(Nimchinsky et al., 1999). Whether these have anything to do with the asylum of the homunculus is, however, highly speculative.

**Acknowledgements** The authors acknowledge helpful suggestions from Jakob Hohwy, Anthony Jack, and two anonymous reviewers. Andreas Roepstorff was supported by a grant from the Danish National Research Foundation to the Center for Functionally Integrative Neuroscience. Chris Frith is supported by the Wellcome Trust and the James S McDonnell Foundation.

## References

- Bateson, G. (1972). *Steps to an ecology of mind*. New York: Ballantine.
- Deacon, T. W. (1997). *The symbolic species. The co-evolution of language and brain*. New York: Norton.
- Deiber, M.-P., Passingham, R. E., Colebatch, J. G., Friston, K. J., Nixon, P. D., & Frackowiak, R. S. J. (1991). Cortical areas and the selection of movement. *Experimental Brain Research*, 84, 393–402.
- Fleck, L. (1979). *Genesis and development of a scientific fact*. Chicago: Chicago University Press.
- Frith, C. (2002). Attention to action and awareness of other minds. *Consciousness and Cognition*, 11, 481–487.
- Frith, C., & Dolan, R. J. (1997). Brain mechanisms associated with top-down processes in perception. *Philosophical Transactions of the Royal Society in London Biological Sciences*, 352, 1221–1230.
- Frith, C. D., Friston, K. J., Liddle, P. F., & Frackowiak, R. S. J. (1991). Willed action and the prefrontal cortex in man: a study with PET. *Proceedings of the Royal Society of London, Series B*, 244, 241–246.
- Gallagher, H., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imagining the intentional stance in a competitive game. *NeuroImage*, 16, 814–821.
- Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Science*, 7, 77–83.
- Jack, A. I., & Roepstorff, A. (2002). Introspection and cognitive brain mapping: from stimulus-response to script-report. *Trends in Cognitive Sciences*, 6, 333–339.
- Jahanshahi, M., Dirnberger, G., Fuller, R., & Frith, C. D. (2000). The role of the dorsolateral prefrontal cortex in random number generation: a study with positron emission tomography. *NeuroImage*, 12, 713–725.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: Chicago University Press.
- Leonard, J. A. (1959). Tactual choice reactions. I. *Quarterly Journal of Experimental Psychology*, 11, 76–83.
- Lhermitte, F. (1986). Human autonomy and the frontal lobes. II. Patient behavior in complex and social situations: the 'environmental dependency syndrome'. *Annals of Neurology*, 19, 335–343.
- McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences USA*, 98, 11832–11835.
- Milner, B. (1963). Effects of brain lesions on card sorting. *Archives of Neurology*, 9, 90–100.
- Monchi, O., Petrides, M., Petre, V., Worsley, K., & Dagher, A. (2001). Wisconsin card sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *Journal of Neuroscience*, 21, 7733–7741.
- Nakahara, K., Hayashi, T., Konishi, S., & Miyashita, Y. (2002). Functional MRI of macaque monkeys performing a cognitive set-shifting task. *Science*, 295, 1532–1536.
- Nathaniel-James, D. A., & Frith, C. D. (2002). The role of the dorsolateral prefrontal cortex: evidence from the effects of contextual constraint in a sentence completion task. *NeuroImage*, 16, 1094–1102.
- Nimchinsky, E. A., Gilissen, E., Allman, J. M., Perl, D. P., Erwin, J. M., & Hof, P. R. (1999). A neuronal morphologic type unique to humans and great apes. *Proceedings of the National Academy of Sciences USA*, 96, 5268–5273.
- Nobre, A. C. (2001). The attentive homunculus: now you see it, now you don't. *Neuroscience and Biobehavioral Reviews*, 25, 477–496.
- Pickering, M., & Garrod, S. (in press) Toward a mechanistic psychology of dialogue. *Behavioral Brain Sciences*.
- Roepstorff, A. (2001). Brains in scanners: an Umwelt of cognitive neuroscience. *Semiotica*, 134, 747–765.
- Roepstorff, A. (2002). Transforming subjects into objectivity. An ethnography of knowledge in a brain imaging laboratory. *FOLK, Journal of the Danish Ethnographic Society*, 44, 145–170.
- Roepstorff, A. (2003). A double dissociation in twentieth century psychology? *Journal of Consciousness Studies*, 10, 62–67.
- Roepstorff, A. (2004). Mapping brain mappers: An Ethnographic Coda. In R. S. Frackowiak et al. (Eds.), *Human brain function* (2nd ed.), 1105–1117. Amsterdam: Elsevier.
- Ruby, P., & Decety, J. (2001) Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nature Neuroscience*, 4, 546–550.
- Sato, Y., Akiyama, E., & Farmer, J. D. (2002). Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences USA*, 99, 4748–4751.
- Stuss, D. T., Levine, B., Alexander, M. P., Hong, J., Palumbo, C., Hamer, L., Murphy, K. J., & Izukawa, D. (2000). Wisconsin card sorting test performance in patients with focal frontal and posterior brain damage: effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia*, 38, 388–402.
- Tomasello, M., & Rakoczy, H. (2003). What makes human cognition unique? From individual to shared individuality. *Mind & Language*, 18, 121–147.
- Tremoulet, P. D., & Feldman, J. (2000). Perception of animacy from the motion of a single object. *Perception*, 29, 943–951.
- Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happe, F., Falkai, P., et al. (2001). Mind reading: neural mechanisms of theory of mind and self-perspective. *NeuroImage*, 14, 170–181.
- Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.