6

Working Memory in Action

This chapter will describe in more detail the positive alternative to the amodal model of reflection critiqued in previous chapters. (This work will continue through Chapter 7.) On the proposed account, the contents of the stream of consciousness in general (and reflection in particular) are constituted by the sensory-based contents of working memory. Our amodal attitudes, in contrast, operate unconsciously in the background, activating, sustaining, and manipulating the contents of working memory. One goal of the present chapter is to provide a sketch of an empirically supported theory of how this system works. Another is to vindicate our intuitive belief that reflective thinking is a form of action (at least in the sense that it is under direct intentional control). The chapter also argues, more controversially, that the seemingly passive nature of much of the stream of consciousness is actually active in nature.

1. Unconscious Goal Pursuit

This section will consider some of the evidence suggesting that goals can be activated, and can guide action, without being conscious. Thereafter the chapter will show how unconscious goals control mental actions of attending and rehearsing, thereby determining the contents of working memory and giving rise to the stream of consciousness.

1.1. Some goals are unconscious

There is a burgeoning literature showing not only that goals can be created in people without their awareness, but also that such goals can motivate action outside of awareness (Dijksterhuis & Aarts, 2010). These findings are important because they demonstrate the reality and efficacy of unconscious goals. The latter play a crucial role in the sensory-based model of reflection and the stream of consciousness, as we will see. Moreover, it should be stressed that the experimental studies in question work with the most demanding possible conception of what it takes for goals to be unconscious, namely, that subjects should have no

knowledge of them. If we equate consciousness with the results of global broadcasting, in contrast (as we have been assuming since Chapter 3), then there will be many mental events that are not globally broadcast (and are hence unconscious) although subjects know, by interpretation, that they occur. Put differently: from the fact that one knows that one has a certain goal one cannot conclude that the goal is a conscious one. For the knowledge in question might be grounded in unconscious inferences from one's own behavior or circumstances, or derived indirectly from sensory-based cues of one sort or another (Carruthers, 2011a). With the bar for unconscious goal pursuit set so high, it is not surprising that it has required extensive experimentation to establish the reality of the phenomenon. But by the same token, once established, we can be confident that it is really quite widespread.

It has been known for some time that unconscious priming can increase the strength of a goal, thereby impacting how someone will act. For example, Bargh et al. (2001) primed the goal of succeeding or doing well by having people complete word puzzles in which terms like "success" figured, whereas a control group completed puzzles in which these terms did not occur. The people primed for success subsequently worked harder and longer at an unrelated task. Likewise, people playing the role of a fishing company who were primed for cooperation did more to replenish the fish stocks than did people who were not so primed. However, this research does not demonstrate the existence of unconscious goals. It only shows that goals can be *influenced* unconsciously. For people in the control groups, too, presumably had the goal of succeeding at the task they had agreed to do, or of cooperating to some degree, and there is no reason to think that these goals were unconscious. (Or none that is provided in these experiments, at any rate. In fact it follows from the sensory-based model of reflection that *all* goals are *always* unconscious, as we will see.) Priming merely increased the *strength* of these goals outside of people's awareness.

There is now, however, an extensive body of evidence showing that goals can be activated, and can influence behavior, outside of people's awareness (Dijksterhuis & Aarts, 2010; Huang & Bargh, 2014). For example, Lau & Passingham (2007) were able to demonstrate that subliminal priming can activate goals that would otherwise not be present. Participants engaged in a task where they either had to make a semantic judgment (indicating whether a word was concrete or abstract) or a phonological one (indicating whether the word was bisyllabic). The judgment that they were to make on each trial was signaled by a prior visible cue (a square or a diamond respectively). Shortly before presentation of the cue, participants were presented with a smaller square or diamond

shape, in conditions where it was rendered either visible or invisible through the timing of a backward mask. They were told to ignore these initial shapes, which could be either congruent or incongruent with the target cues. What the experiment found was that when the prime was both subliminal and incongruent with the target cue, people made more errors and were slower to respond, suggesting that the prime had activated the opposite goal, which then conflicted with the cued goal. (There was no effect when the prime was visible.) Moreover, brain imaging revealed increased activity in cortical areas associated with semantic and phonological judgments respectively that was consistent with the subliminal prime, as well as increased activity in dorsolateral prefrontal cortex in cases where the subliminal prime and target cue were in conflict. These findings strongly suggest that the goal of making a semantic judgment, for example, was unconsciously activated by the subliminal prime and that it interfered with the operation of the consciously caused goal of making a phonological judgment.

Similarly, Marien et al. (2012) designed a number of experiments to investigate the question whether subliminally primed goals would interfere with tasks that require significant executive-function resources. For unconscious goals, if they exist, should compete for control of attentional resources in the service of their own achievement, just as do consciously caused ones. This would then reduce the resources available to serve other goals, such as the ones that participants are pursuing consciously. In some of the experiments the participants' primary task was, in effect, a working-memory one. They were shown a display of four letters and then, after a delay of a few seconds, were presented with a probe letter and required to judge whether or not it belonged to the initial set. Tasks of this sort are known to require inhibition of memories deriving from immediately preceding displays, which would otherwise interfere with judgments relating to the target. In one experiment participants were subliminally primed with the goal of socializing, whereas in another they were primed with a goal known to be important to them individually from a previous questionnaire. Reaction times in trials where the working-memory task required inhibition of previous memories were significantly slower among subliminally primed participants than in controls, suggesting that their newly active unconscious goal had hijacked some of the attentional resources required for the task. Marien et al. were able to show, moreover, that this did not just result from a decrease in motivation toward the primary task, since the effect was not moderated when people were paid to succeed in that task. The finding also generalized to other tasks requiring attentional control, such as detecting errors in text.

1.2. All goals are unconscious

These and voluminous other data demonstrate that goals *can* (and sometimes do) operate unconsciously. This secures one of the main assumptions required for the truth of the sensory-based accounts of reflection and the stream of consciousness to be presented in this and the following chapter. But of course such data cannot demonstrate that goals *always* operate unconsciously. Nevertheless, this stronger conclusion is warranted, I suggest, by combining the claim that goals *can* operate unconsciously with three additional assumptions. The first is that goals themselves are amodal states, as discussed in Chapter 2.2. (Recall that goals are non-affective intention-like states, which have propositional contents structured out of amodal concepts.) The second assumption is the global broadcasting account of consciousness, discussed in Chapter 3.2. (Recall that access-conscious states are those that are made widely available to other states and faculties of the mind. Moreover, since one can know through swift self-interpretation that a mental state is occurring, it is not sufficient for a state to be conscious that one should know intuitively that it exists.) And the third assumption is the denial of an amodal global workspace, defended in Chapters 4 and 5. (Recall that the evidence suggests that amodal attitudes cannot be globally broadcast, and that amodal concepts can only be conscious when bound into the contents of globally broadcast sensory-like states.) When taken together, these claims entail that no goals are ever conscious. Hence they can *only* operate unconsciously.

This argument is not conclusive, of course. In part this is because it depends on a number of inferences to the best explanation. But it is also because we have yet to consider the possibility that some kinds of conscious sensory-based event (such as saying to oneself, in inner or outer speech, "I will be a home-owner one day") can be, or can *constitute*, a sort of goal-state. Some versions of this idea will be considered in Chapter 7.

2. Attention as Action

Recall from Chapter 1.1 that conscious reflection is the subset of the stream of consciousness that strikes people as being under their control. One can, to a significant degree, *direct* the course of one's reflections in the service of one's goals. One can focus on thinking about the solution to a particular problem, or one can choose to replay a familiar fantasy in order to relax before falling asleep. The active nature of reflection is something that I endorse. But if reflection is sensory based, as I suggest, and involves rehearsal and manipulation of sensory-based information in working memory, then attention, too, must be under intentional

control. For as we have seen, attention is the primary cause of global broadcasting, and is a major determinant of the contents of working memory. The present section will elaborate and begin to defend this idea.

2.1. What are actions?

Before discussing the active nature of attention, a prior question to consider is what actions, in general, are. From a common-sense perspective, the paradigm case of an action is a bodily movement that is caused and controlled by an immediately preceding decision to act. First we engage in practical reasoning in light of our beliefs and desires; then we decide; and the decision (if the decision is for the here-and-now) creates an intention that causes and controls the motor processes that issue in the movements decided upon. But of course practical reasoning often concerns what to do in the future, rather than the here-and-now. In this case the result of a decision is the formation of a standing intention, or intention for the future, which is later activated when the circumstances required for the execution of the action obtain.

It seems plausible, however, that these two cases follow essentially the same pattern. That is, in the second case the activated intention interacts with judgments about the current circumstances to issue in a *decision* to act *now* (but without any need for further practical reasoning about whether or not to act at all; Bratman, 1987), which leads the intention to initiate and control the resulting bodily movements. So the suggestion that actions are bodily movements caused by immediately preceding decisions can be allowed to stand. If the idea of *mental* action is to make any sense, then this analysis needs to be made more abstract, of course. One could say, perhaps, that an action is any *event* (whether involving bodily movements or not) that is directly caused and controlled by an immediately preceding decision.

Although intuitive, we now know that not all bodily actions fit this pattern. This is because many movements are caused and controlled directly by perceptions of affordances in the environment in the *absence* of a decision *not* to act. Perceptions of familiar tools, for example, automatically activate the motor plans for their use, which need to be inhibited by top-down executive signals if they are not to be carried through to completion (Frith et al., 2000; Negri et al., 2007). As a result, people with certain forms of frontal-lobe damage can suffer from *utilization syndrome*, meaning that they cannot inhibit themselves from grasping and using things in accordance with those things' affordances (Lhermitte, 1983, 1986). For example, if a glass is placed on the desk, the patient will reach for it; and if a pitcher of water is placed on the desk as well, he will pour water into the glass and drink, irrespective of the context. Intuitively these are nevertheless

forms of action. They are controlled and apparently goal-directed, and lack the ballistic quality of obvious non-actions like the knee-jerk reflex.

One option would be to say that there are two distinct kinds of action. There are movements that are caused and controlled by immediately prior decisions. (These might be called "intentional actions.") And there are movements that are caused and controlled by motor plans in the absence of a decision not to act. (These might be called "automatic actions.") Another option, however, would be to extract the common core from the two kinds, and to say that an action is a bodily movement that is caused and controlled by a motor plan. For in the case of intentional actions, too, the result of a decision to act is presumably the activation and execution of an appropriate set of motor schemata. If this latter option were to be adopted, then that would probably restrict the range of potential mental actions to the set of mental events (such as inner speech) that are caused and controlled by mental *rehearsals* of action, which likewise activate an appropriate set of motor schemata. Attentional processes would thus likely be excluded.

It is unclear which of these alternatives better reflects the underlying reality of the mind, nor which will prove more theoretically fruitful for cognitive science. But even if it should turn out that the motor-plan conception of action is more scientifically robust, we can still say that mental events that are caused and controlled by an immediately preceding decision are at least action *like*. That will be sufficient for our purposes.

2.2. The control of attention

Attention is a form of basic action (or is at least an action-like event), I suggest. Attending is something one can "just do," without knowing how one does it, just as lifting one's arm is something one can do without knowing how. Attention is not, of course, a physical action. Rather, it is a mental action. What makes it active is that it is (generally) caused by *decisions* to attend, where these decisions are informed by knowledge of the context together with current goals and standing values. This account will be supported in due course, in part through its capacity to explain familiar features of conscious reflection and the stream of consciousness more generally. (Sections 3 and 4 of this chapter will focus on reflection, while Section 5 discusses mind wandering.) It is also supported by recent models that see the role of anterior cingulate cortex as mediating between bottom-up and top-down forms of attention, making cost–benefit calculations about the allocation of top-down executive resources (Shenhav et al., 2013). But at least a limited version of the idea comports quite well with aspects of common-sense belief. For we know that one can, at will, comply with a request to attend to

one thing rather than another (just as one can comply with a request to raise an arm), and we know that one can *try*, but fail, to attend to something (just as one can try, but fail, to raise an arm). In these respects, at least, attending seems thoroughly action-like.

Often, of course, things just seem to *grab* one's attention. In such cases attending seems passive rather than active. It seems to be something that happens to one rather than something that one does. But most instances of bottom-up attentional capture are really active in nature. Or so I will initially suggest here, and then defend more fully in Section 5. (Exceptions might include the impact of highly salient stimuli like loud noises, bright flashes, or sudden dramatic movements, which can redirect attention without much involvement from frontal decision-making systems.) As we noted in Chapter 3.4, the bottom-up cortical attentional system continually monitors unattended aspects of the input, checking those representations for relevance to ongoing goals and existing values, and competing to redirect the top-down attentional system accordingly (Corbetta & Shulman, 2002; Corbetta et al., 2008). When it succeeds, and attention is captured, this is best seen as resulting from a *decision* to attend to the novel stimulus because of its apparent relevance. The difference is just that in such cases one is generally unaware of the rationale underlying the decision. Hence such cases do not strike one intuitively as agentive in nature, in contrast with cases where, for example, one knows that one is actively complying with an experimenter's request to attend to one thing rather than another.

We noted in Chapter 2.1 that value-processing interacts with perceptual processes at many different levels within perceptual networks (Barrett & Bar, 2009). One way of understanding why this should be so is that it drives the competition for attentional resources. One cannot attend to everything, of course. Indeed, quite the contrary: attention is a highly limited resource (Cowan, 2001). Hence it is important that it should be allocated wisely, directed toward those stimuli that matter most. And just as one might then expect, the bottom-up attentional or "salience" network is partly defined through the involvement of sub-cortical valuational systems (Menon, 2011; Yeo et al., 2011). Perceptual processing thus involves a hierarchy of increasingly fine-grained, context-sensitive value judgments, with the eventual winners becoming the targets of the sort of top-down attention that is needed for them to be globally broadcast, and thus to become widely accessible to numerous systems throughout the brain. Just as actions are generally initiated by decisions made in the light of value judgments together with current circumstances, so, too, are allocations of top-down attention, I will suggest.

2.3. Unconscious decisions

Why should we regard the events that issue in the control and redirection of attention as *decisions*, however? Primarily, this is because those events fit much of the functional profile of decisions, and so are at least decision-*like*. We noted in Chapter 2.2 that decisions are events that conclude an episode of practical reasoning about what to do, immediately issuing in intentions that cause and control the decided-upon actions. And it seems that attentional shifts, likewise, are caused by events that conclude episodes of (unconscious) reasoning about which of the representations competing for attentional resources are the most relevant. If we assume that such shifts are themselves actions, then the event that concludes the process of reasoning about what to attend to, which causes such a shift, looks much like a decision. And indeed, it seems to be the very same sorts of computation of the expected value of cognitive control, taking place in anterior cingulate cortex (which also serves as the primary mediator between bottom-up and top-down forms of attention), that issue not only in actions of a physical sort but also in redirections of attention (Shenhav et al., 2013). So there are theoretical gains to be had from treating both sorts of conflict-resolving event as kinds of decision. We can thereby unify two sets of phenomena under the same explanatory umbrella, constituting a single natural kind.

It might be objected that the events that shift or maintain attention are not (or not generally) *personal level* decisions. It is not *the person* who decides to shift attention to the sound of his own name, and away from a current conversation at a party, thereby giving rise to the so-called "cocktail party effect." Rather, it is a subcomponent or module within the person that does so. Now, sometimes the language of "personal" versus "subpersonal" is used just to mean just "conscious" versus "unconscious." But in that case an argument exactly parallel to the one presented in Section 1.2 for the case of goals will establish that there are no such things as personal-level (conscious) decisions. At other times personal-level mental events are understood to be those that the entire agent controls, or that all (or most) of the mental states of the agent can contribute to. This idea can still be made good sense of, as we will see in Chapter 9. But it does nothing to challenge the claim that the events that determine shifts of attention are a species of decision. Let me elaborate.

In denying that decisions themselves are ever conscious one need not deny that any of the processes that *issue in* decisions are ever conscious, of course. Nor need one deny that there are any personal-level decisions in the above whole-agent sense. So consider a case where one reasons, consciously, about whether to shift attention for a while to another conversation. When the sound of one's own

name pops out, one might say to oneself in inner speech, "She is talking about me. Should I listen in?" The resulting contents are globally broadcast to many different regions of the mind, giving all of them a chance to contribute to the decision-making process. (The process that shifted attention to the sound of one's own name in the first place, in contrast, was a much more local affair.) But any representations evoked by one's conscious reflections will still need to compete with one another to influence one's attentional control processes. And it is still likely to be cost–benefit analyses conducted in anterior cingulate that determine the result. I suggest, in fact, that both conscious and unconscious decision-making processes issue in events of the same kind: unconscious decisions.

2.4. How remembering works

The idea that attending is action-like can be put to work in understanding how, in outline at least, memory-search operates. (Recall from Chapter 4.1 that episodic memory is heavily attention-dependent.) And note that memory-search, too, is intuitively active in nature, at least on many occasions. (Exceptions will include cases where memories seem to spring to mind unbidden, perhaps evoked by a spoken phrase or a familiar scent. But again I will argue that even these seemingly passive instances of remembering are really action-like in nature.) One can try but fail to remember something. And one engages in active, motivated, memory search every time one tries to answer a question. Moreover, remembering can unfold over time in accordance with one's goals, as one gradually builds and elaborates a memory of some important event. Yet we know that conscious remembering results from targeted attention (De Brigard, 2012). So if remembering is active in nature, this reinforces the claim that attending is too.

It has been known for some time that parietal cortex plays important roles in episodic remembering (Wagner et al., 2005). Indeed, damage to ventral parietal cortex causes a form of *memory neglect* (Berryhill, 2012). People with damage to this region have trouble spontaneously activating relevant details into memory, whereas they can still recover those details when directly questioned (Berryhill et al., 2007). These findings make sense once we realize that episodic memory involves both top-down and bottom-up attentional networks, which are centered on the intraparietal sulcus and areas of ventral parietal cortex respectively (Cabeza, 2008; Ciaramelli et al., 2008). When searching for a memory one uses the cues provided to evoke activity in those sensory-involving regions where the information is stored, utilizing the back-projecting conceptual-to-perceptual networks discussed in Chapter 3.5, and using the top-down attentional system to target and globally broadcast the results.

A prediction of this account is that activity around the intraparietal sulcus should be highest when memory search is most effortful and attentionally demanding, and thus when memory performance and memory confidence are both low. That is, indeed, what we find (Kim & Cabeza, 2007). When memories are *evoked*, in contrast, the bottom-up attentional system monitors the representations produced and evaluates their relevance to one another and to the probe, attracting top-down attention when there is a good enough fit. This predicts that activity in ventral parietal cortex should be highest when memories are rich in detail, and when both confidence and accuracy are highest. This, too, is what we find (Kim & Cabeza, 2007).

This model of the involvement of the two attentional systems in episodic memory is also consistent with the finding of strong neural connections between ventral parietal cortex and the hippocampal formation, which is known to play a vital role in memory storage and retrieval (Vincent et al., 2006). It makes sense that evoked information should be received by ventral parietal cortex in the first instance (just as bottom-up perceptual contents are), so that they can be evaluated for relevance, attracting top-down attention (and hence becoming globally broadcast) if they should pass muster.

While evidence of the involvement of top-down attention in episodic remembering is robust (Hutchinson et al., 2009; Cabeza et al., 2010; De Brigard, 2012), there may be less overlap in the bottom-up saliency mechanisms involved in perception and episodic memory respectively. Specifically, while the bottom-up perceptual-attention system is largely lateralized to the right hemisphere, the bottom-up memory system is mostly lateralized to the left (Ciaramelli et al., 2008; Hutchinson et al., 2009). Moreover, the regions of ventral parietal cortex that are involved are only partly overlapping, with the perceptual-saliency system being located around and above the temporo-parietal junction, whereas the mnemonic-saliency system is lower, overlapping parts of the temporal lobe (Sestieri et al., 2010; Cabeza et al., 2012). It makes sense that the mechanisms that process the relevance of perceptual and mnemonic representations should be to some degree specialized, and also that they should be partly segregated so that both processes can continue in parallel. For one needs to monitor the environment for relevance while engaged in stimulus-independent thought, of course. And we have reason to think that activated memories, too, are continually monitored while one focusses on an external task, causing one's mind to wander when some of those memories are deemed to be sufficiently relevant.

On the account sketched here, evoked memories are likely to be just as active as memories that are searched for, although they are less obviously so from the perspective of the person doing the remembering. (In the former case a memory

just appears in consciousness, seemingly unbidden; whereas in the latter case the memory will have been preceded by a question of some sort that one is aware of.) Consider a case where a waft of cinnamon when passing a shop evokes a sudden vivid memory of a dinner eaten in the Moroccan quarter of the French city of Avignon some years before. The smell in question activates associated representations that are linked both to it and to one another as a result of the original event. These are received in ventral parietal cortex where they are processed, and from where messages are exchanged with ventrolateral prefrontal cortex and subcortical value systems, assessing their relevance to one's goals and values. Since the Avignon dinner is linked to a number of important values (a shared experience with one's spouse, love of all things French, and so on) it is found relevant enough to enter into competition (via the anterior cingulate) with the goals controlling the current focus of attention. By hypothesis, it wins this competition, resulting in a decision to redirect top-down attention to the memory representations in question, rendering them conscious. The experience of the memory thus results from an action or action-like event (a redirecting of attention), although it doesn't seem like that to the agent.

2.5. Memory search and spatial search

Not only is attention action-like in the ways discussed above, but it also plausibly evolved from related forms of *physical* action. Recall from Chapter 3.4 that the frontal eye-fields are an important component of the top-down attentional network. These regions of the dorsal prefrontal cortex are parts of premotor and motor cortex, and play a central role in the control of eye movements (as does the intraparietal sulcus). And we know that eye movements are generally under intentional control, even in infancy (Kidd et al., 2012). Often, of course, one's eyes and one's covert attention will shift together; and generally one attends to what one is overtly looking at. But we know that this is not strictly necessary. One can look in one direction while attending elsewhere. And many eye movements are best thought of as exploring a single object of attention, without being accompanied by any shifts in attention. Moreover, attention can zoom in or out even when one's eyes and one's attention are both focused in the same direction (Kosslyn, 1994). So the frontal eye-fields appear to form a common component in two partially distinct controllers: one for overt eye movements and one for covert attention.

As we will see in Chapter 8, attentional systems are highly conserved across species. And we know that overt and covert visual attention can be deployed separately in both birds and mammals (Mysore & Knudsen, 2013). Moreover, it is plausible that capacities for covert attention evolved from earlier capacities for

the control of eye movements. This would explain why they are subserved, in part, by the same brain regions. But in any case what surely follows is that the mechanisms that control top-down covert attention are ideally positioned to be intentionally controlled by one's goals and decisions. For eye movements, like most other forms of movement, are under intentional control. So the neural wiring necessary to exert that control would already have been in place when the frontal eye-fields (or their evolutionary precursors) began also to direct covert forms of attention.

Interestingly, a number of recent studies support both the active nature of attending and the suggestion that covert attentional control may have evolved from earlier mechanisms for controlling overt movement. These studies demonstrate commonalities among spatial search, attentional search of the environment, and attentional search of memory, together with evidence of a phylogenetic progression from the former to the latter (Hills, 2006). Not only is there evidence of common search strategies in the three domains, but also evidence that the neural mechanisms underlying these strategies in all three domains are similar, involving reward-processing doperminergic networks (Hills & Dukas, 2012).[1]

All animals confront the problem of searching for valued resources, which involves a trade-off between exploration and exploitation. Often such resources are patchily distributed in the environment, and patches differ in their size and value. The problem for any animal is then: when to stay and when to go? Time spent exploring is time that is not spent exploiting whatever valuable resource is at stake. But on the other hand, if one spends too long extracting value from any given patch, one may lose out on more valuable patches nearby. In the context of foraging, this problem has been extensively studied, and many of the search heuristics employed by animals in different environments—most of which approximate to an optimal foraging strategy—are known (Stephens & Krebs, 1987).

More recently, people have realized that *cognitive* search shares much of the same structure. Not everything in one's immediate environment is equally interesting, and features of interest are often patchily distributed. The same holds for regions of one's visual field. Both overt and covert attention thus face the same exploration–exploitation trade-off (Hills & Dukas, 2012). Moreover,

[1] Moreover—and highly relevant to the topic of the present book—variations in capacities for dopermine synthesis in an important part of this network in humans—the ventral striatum—correlate with individual differences in working-memory capacity (Cools et al., 2008). It makes good sense that well-applied strategies for when and why to shift one's attention should be an important component of people's working-memory abilities.

many forms of memory, too, are patchily distributed. Words are linked to one another in clusters, for example, with names of common farm animals being semantically linked to one another while only being distantly linked to names of common African animals, which are in turn strongly linked to one another. Experiments suggest that when people are asked to generate names of animals they follow essentially the same strategy that creatures use when exploiting a patchy environment: they stick with farm animals until names are no longer easy to retrieve, then switch to African animals, and so on (Hills et al., 2012).

These commonalities among spatial search, attentional search, and memory search do not demonstrate that the same mechanisms are involved in each, of course. But they do suggest that each is equally active in nature, at least to the extent that each can be controlled by the same heuristics that control behavioral search of the physical environment. And as we also noted, this conclusion is further supported by the existence of shared decision-sensitive neural mechanisms (especially the frontal eye-fields).

2.6. Habits of attention and cross-cultural differences

If attention is a form of action, as I have been suggesting, then it should be controllable in any of the ways that overt action can be controlled. My focus in this section so far has been on *intentional* control, arguing that attention, too, is controlled by (unconscious) decisions taken in the light of one's goals and standing values. But actions can also be *habitual*, of course. In such cases features of the context will trigger a motor schema directly unless inhibited, without the need for evaluation or decision making (Lisman & Sternberg, 2013).

It is certainly intuitive that patterns of attention can become habitual, although I know of no direct scientific work on the topic. Someone might say, for example, "Whenever I enter a new place, I find myself scanning for the exits." Cases like this could just as well be explained in terms of intentional attention-allocation, however, caused by persisting motivational states (in this case, anxiety). More promising, perhaps, are cases of habitual *thought*, given that these are caused by the way in which one directs one's attention. Someone who habitually calls up the same set of images when lying down to sleep at night may well exhibit patterns of attention that are genuinely habitual and non-intentional.

Indirect evidence of habitual use of attention can be gleaned from the literature on cross-cultural differences in cognition, however, particularly the finding of systematic use of "holistic" versus "analytical" processing styles in Eastern versus Western cultures (Nisbett, 2003). Among these differences are variations in the way that members of the two cultures attend to, and hence recall, visual scenes

(Ji et al., 2000; Masuda & Nisbett, 2001). For example, when presented with a picture of some fish swimming, Japanese people might pay as much attention to the background of reeds and rocks as they do to the fish themselves. As a result, when shown a picture of a lone fish and asked to recall whether it was in the original picture, they are helped if contextual cues are provided. In contrast, Americans shown the same picture will attend mostly to the focal fish, paying little attention to the background. As a result, contextual details fail to aid them when asked to recall an individual fish. Indeed, they are *better* able to recognize an individual fish in the *absence* of any background.

How are these findings to be explained? We know that the differences don't result from long-term structural cognitive–perceptual differences between Easterners and Westerners. In part this is because the effects are reversible using simple forms of priming (Oyserman & Lee, 2008). A Westerner who is asked to read a passage containing repeated use of the terms "we," "us," and "our" (emphasizing collections of people), will thereafter perceive and recall the pictures in a manner normal for an Easterner. Likewise, Japanese people who read the same passage only with the words "I," "me," and "mine" substituted throughout (emphasizing the individual), will thereafter perceive and recall the pictures in the manner of a Westerner. It also seems unlikely that these patterns of attention are *motivated* ones, with Easterners frequently taking (unconscious) *decisions* to attend to the background while Westerners often decide to ignore the background. Rather, what seems most plausible is that while both patterns of attending remain easily *available* to members of both groups, differences in cultural norms and expectations give rise to an *habitual* tendency on the part of Easterners to pay attention to contexts and situational factors, as well as to a corresponding habitual tendency on the part of Westerners to direct attention towards focal individuals. If this is true, then it provides an additional reason to think that attending is genuinely action-like.

3. Mental Rehearsal and Mental Manipulation

Recall from Chapter 4.4 that the functions of working memory are normally thought to include capacities to *activate*, *sustain*, *rehearse*, and *manipulate* representations. The first two functions depend directly on the allocation of attention. Section 2 has argued that this is an active process, and is generally directed by decisions taken in the light of one's goals and values. The present section will argue that rehearsal and manipulation are also forms of action. In addition, it will begin to demonstrate how the present framework can explain familiar features of reflection.

3.1. Mental rehearsal as off-line action

As we saw in Chapter 4.4, mental rehearsal is really just off-line action, with the predicted sensory consequences of action attended to and globally broadcast while movements of the muscles are suppressed (see Figure 5). So mental rehearsal can be guided and influenced by anything that can guide and influence action generally (with the exception of afferent sensory feedback from the body or world, of course, since no movements are really made). In particular, like overt action, mental rehearsal can be guided by current goals and values, as well as other items of information that are active in one's mind at the time. Our focus in this section will be on rehearsal of non-speech actions, together with their role in prospection and future decision making. Inner speech will form the topic of Section 4.

Mentally rehearsed actions cannot be motivated in exactly the same way that overt actions are, of course, if only because something must lead to the suppression of overt movement in the former cases. Indeed, part of what it means to be a reflective (as opposed to an unreflective) person is that one frequently inhibits one's initial overt response to a situation or question and mentally rehearses it instead, evaluating the imagined result. Such inhibition might be habitual, or it might be motivated by chronic caution or some other goal or value. In effect, one thinks, "I am tempted to say/do X. But is that really the best response?" In such cases the goals and values that underlie the mental rehearsal can be just the same as those that would motivate the overt action, but perhaps with the added goal of not acting prematurely. (If the person's reflectiveness is *habitual*, in contrast, then the motivational factors involved can be exactly the same.) The result is that the decision is a decision to activate-and-suppress (that is, mentally rehearse), rather than to act.

Not all mental rehearsal is externally prompted, of course. Indeed, quite the contrary. Since so-called "stimulus-independent thought" occupies much of our waking lives, and since a good deal of this involves mental rehearsals of action of one sort or another, much mental rehearsal must involve more than mere inhibition of an externally prompted action. (This is a topic we will return to in Section 5.) But one plausible possibility is that what it *is* to be engaged in, or to switch into, stimulus-independent thought is that one has (or that one activates) the standing goal of inhibiting the execution of any intentionally controlled action.[2]

[2] Of course one can engage in stimulus-independent thought while walking or driving, say. But these actions are very likely habitual ones, involving interactions between perception and the motor system, without the intervention of goals or decisions.

Consider, for simplicity, some cases where reflection is already ongoing. Suppose a practiced skier is engaged in a fantasy ski-run down a mountain she will visit that weekend. In the course of this she imagines a tree straight ahead with some rocks to its left. (This might be thrown up by attention directed at memories of the mountain, or by quasi-random activation and attention to things that one might confront while skiing. See Section 5.) As a result, she immediately activates the motor schemata for a right-hand turn, and her imagined ski-run unfolds accordingly. Here the rehearsed action is likely to be habitual, prompted by the imagined scene alone.

Alternatively, think of someone engaged in a fantasy about a Caribbean holiday. He is sitting at a bar with a bottle of his favorite beer in front of him and with the beach behind him. The beer is seen as positively valenced, which activates the motor schemata for reaching out to grasp, lift, and drink it. Accordingly, that is what he imagines himself doing. Here what motivates the imagined action is just what might motivate the real action: the seeming-goodness of drinking the beer, combined with a perceptual representation of it as being within reach.

No doubt there is much more that might be said. (And some of it *will* be said in what follows.) But enough should have been done to support the thesis of this subsection. This is that mentally rehearsed actions are actions in pretty much the same sense as overt ones. They involve activations of motor schemata that are either immediately prompted as part of a habit, or that result from the same sorts of decision-making processes that issue in overt action.

3.2. Three kinds of dynamic imagery

The imagery that results from mental rehearsal of action is, of course, dynamic in character. It will be imagery of perceived or felt movements, together with their likely consequences. But there are other ways of generating dynamic imagery. One is from memory of perceived movements. Creating an image of a horse galloping across the landscape, for example, will require a guided search of memory, using top-down projections of the concept HORSE GALLOPING in coordination with targeted attention. It probably works somewhat as follows. Activation of the concept HORSE GALLOPING triggers associated representations stored in a distributed manner across regions of visual cortex. Those that match the concept sufficiently well become targets of attention, and the resulting dynamic image is thereby made globally accessible.

A similar distinction can be made in connection with dynamic auditory imagery. Imagery of musical melodies, for example, can be created through off-line rehearsal of the actions that would issue in those melodies, by either singing

or playing a musical instrument. If the melody is a familiar one, it will be recalled by activating a well-rehearsed sequence of motor schemata. In this case memory for the sound-sequence is really a form of motor-memory, with mental rehearsal of the actions that would produce that sequence being used to create appropriate auditory images. If the melody is a novel one, it will be generated by constrained quasi-random activations of a sequence of component motor schemata. (See Section 5.) On the other hand, a melody can be imagined by directing attention at a memory of the relevant sequence of sounds, perhaps using some sort of conceptual trigger. ("The opening bars of Beethoven's Ninth.") And it may well be possible to create a novel melody in imagination in similar fashion, by targeting memories of short phrases or individual notes in a constrained but quasi-random manner. (Again, see Section 5.) In the first sort of case the active nature of auditory imagery is obvious, since it results from mental rehearsals of sound-producing actions. But even the second sort of case is no less active in nature, given that directing and shifting attention qualify as forms of mental action, as was suggested in Section 2.

There is, in addition, a third way in which dynamic imagery can be produced. This is by directing rehearsed motor movements at an existing perceived object, thereby transforming one's image of the latter in the manner that might be predicted if one were really acting in that way on the object represented. This sort of mental manipulation of imagery has been extensively studied, especially using mental rotation experiments. As a result, we know that activity in motor and premotor cortex is reliably observed during mental rotation (Richter et al., 2000; Vingerhoets et al., 2002). Moreover, mental rotation is impacted by concurrent movement. For example, if one moves one's hands to rotate a real object while mentally rotating an image in the same direction, then the latter is speeded up, whereas overt rotation in the contrary direction slows it down (Wexler et al., 1998). In addition, applying transcranial magnetic stimulation (TMS) to the hand region of primary motor cortex interferes with mental rotation (Ganis et al., 2000).[3]

In all of these cases it should be emphasized that dynamic imagery is importantly predictive in nature. For as representations of the target movements or changes are globally broadcast, they activate both memories and predictive

[3] There are other data that suggest, however, that mental rotation can be conducted by infants who as yet lack the physical coordination to really rotate anything. Thus Moore & Johnson (2011) found that three-month-old male—but not female—infants could recognize the completion of a previously habituated partial rotation, in contrast with its mirror-image rotation. One possibility is that there are innate linkages between structures in motor cortex and visual cortex, which mature in males in advance of capacities for overt motor control.

inference-mechanisms to create representations of the likely immediate consequences of those movements or changes. If attended to these, too, will be globally broadcast. Thus when one visually imagines a galloping horse one will be apt also to hear the sound of its hoof-beats in auditory imagination. And when one rehearses the action of throwing a stone at a window, one will likely form an image of the trajectory of the stone and the subsequent shattered glass. As we will see in due course, the predictive character of imagery is one of the things that makes it useful, enabling us to discern and evaluate the likely consequences of our actions in advance.

3.3. Mental manipulation

In the literature on working memory, "manipulation" has both a narrower and a wider meaning. In the narrower sense it refers to the capacity to bring about changes in one's imagery by activating manual motor schemata, as in the mental rotation experiments described above. But in the wider sense, it refers to the capacity to construct ordered sequences of images in the service of one's goals. Examples would include mental arithmetic, spatial planning, and prospective reasoning. (The latter will be discussed in Section 3.4.) In this wider sense mental manipulation is a form of skillful, knowledgeable, intelligent action, and is motivated accordingly. It is this that will be our focus here.

Consider someone tasked with subtracting 17 from 32 in her head, for example. What ensues will be a controlled sequence of auditory imagery, visual imagery, or both combined, implementing well-rehearsed procedures. She might picture the two numerals one above the other, for instance, as if written on the page, and then imagine removing a unit from the "3" and placing it in front of the "2," before saying to herself, "Seven from twelve is five," thereafter mentally inscribing "5" beneath the "7." She then transfers her attention to focus on the left column, which now contains a "2" inscribed above the "1." This immediately enables her to construct the answer: 15. This sequence is plainly active, involving mental rehearsals of familiar actions and action-sequences, combined in a flexible way in response to the initial problem or goal.

Now consider someone driving home after work, who hears on the radio that her normal route has been blocked by an accident. Supposing that an alternative route does not immediately come to mind following a search of memory, she now needs to engage in a sequence of spatial planning. She might call to mind a memory of a road map of the area, for example, or some memory images of major nearby landmarks or roads, which come already flagged as "places I could get to from here" (perhaps by swiftly entertained imagery of the intervening route). She might then say to herself, "From Georgia I could go down Wayne,"

but this prompts a memory image of long queues at the traffic lights at the junction of Wayne and Flower. This motivates an additional search of memory. Returning to the initial partial map of the nearby roads, for example, she might imagine herself driving up New Hampshire Avenue. As the predicted sequence of sights (drawn from memory) flips by, she says to herself, "Then I could go up Sligo." The resulting image of the initial part of Sligo Parkway feels highly familiar, or issues in a swift sequence of images that terminates at home. So she turns toward a road she knows will lead her to New Hampshire Avenue.

Notice the close parallels between this sequence of working-memory contents and what might happen in a case of real, overt, spatial exploration.[4] The same goal in each case motivates searches of memory, and selects and activates action schemata guided by those memories. In a case of real travel, of course, a sequence of perceptions is caused by one's real movement through the world, whereas in the case of spatial planning the images are called up predictively from memory following each rehearsed action-schema. But both are equally active and intentionally controlled.

Notice, too, how there is a place for something resembling real discovery in cases where one manipulates working-memory contents for purposes of planning. Had our agent *really* driven up Georgia and down Wayne, she might really have encountered a long line of traffic as a result. This is a new (although perhaps predictable) item of information about the world. When rehearsing the route, in contrast, she cannot discover anything that she doesn't already know (albeit implicitly), of course. But what she *can* do is evoke knowledge that would otherwise have remained inaccessible. It is only by imagining herself driving toward the junction of Wayne and Flower that she activates the knowledge that the approach roads are often snarled with traffic. She thereby discovers (in the sense of bringing to consciousness) that this is not a good route to take.

What we have found in these examples extends to mental manipulation of contents in working memory quite generally. Goals motivate both searches of memory and mental rehearsals of action. The resulting contents then evoke yet other memories, or predictions grounded in existing knowledge or procedures. And at each stage (if the account sketched in Section 2 is on the right lines) a multitude of evoked contents will compete with one another through bottom-up attentional networks to attract the spotlight of top-down attention and enter the stream of consciousness. Generally the contents deemed to be most relevant to

[4] The main difference is that had she really driven up Georgia and then down Wayne, only to find her route blocked, she would probably not have needed to return to her point of origin in order to find her way home.

the overall goal or goals will win. The result is that not only are the individual stages of the working-memory sequence active in nature, but so too is the sequence as a whole.

3.4. Prospection

Prospective reasoning, in the most general sense, is reasoning about the future. But in the recent literature it has come to acquire a narrower meaning, restricted to cases where one *imagines* or envisages oneself performing one or more future actions, generally for purposes of decision making (Gilbert & Wilson, 2007). It is now widely accepted that capacities for prospection and episodic memory are tightly linked (Schacter et al., 2007; Buckner, 2010). The hippocampus, in particular, is heavily implicated in both; and amnesic patients turn out to have mirroring difficulties in envisaging their own futures. Moreover, some have claimed that the resulting capacity for "mental time-travel" is uniquely human—indeed, that it is perhaps *the* major cognitive factor that distinguishes humans from other animals (Suddendorf & Corballis, 2007; Suddendorf et al., 2009). These claims about human uniqueness will be examined—and critiqued—in Chapter 8.

Episodes of prospective reasoning are often prompted by an externally presented question or choice, as when one is asked whether one would rather do A or B, or when one receives a job offer and has to decide whether or not to accept. But they are also frequently self-initiated during mind wandering. Someone entertaining a job offer, for example, will probably revisit the choice on a number of occasions each day for multiple days or weeks before reaching a decision, often breaking off from other attentionally demanding activities to do so. Internally prompted forms of prospection will be considered in Section 5. Here we can focus on cases where the initial stimulus comes from outside.

Sometimes prospective reasoning will start from mental rehearsals of the actions being considered. In such cases the process is quite obviously active in nature. If asked, "Would you rather climb that fence or walk around to the gate?," one may rehearse the two actions involved. In the one case this issues in imagery of one approaching the fence, clambering up it and then down the other side. As a result one experiences, predictively, some of the effort likely to be involved, and one's imagination may also become elaborated with likely side-effects, such as a painful fall to the ground if one catches one's leg on the top rung. (Such elaboration is especially likely to occur if episodic memories of oneself or others suffering similar falls are evoked.) In the other case mental rehearsal will issue in imagery of walking some distance up the hill to the gate that can be seen in the distance (or that is known to be there), again resulting in

a prediction of the effort involved. As we will see shortly, the decision that one takes between the two options will often be grounded in one's overall affective reaction to each scenario.

On other occasions prospection can start from imagery of some of the known consequences of the action being considered. Confronted with a job offer that would require relocating to Chicago, for example, one may imagine being in Chicago. This is likely to build from episodic memories of visits one has made to the city (where available), or from images that one has of it from movies or the TV news. One might imagine walking along the lake shore with the city skyline alongside; or one might imagine taking a boat ride on one of the canals. Or one might simply imagine oneself being in one of Chicago's streets. Again, what one imagines is likely to become predictively enriched as further knowledge of the city or the job opportunity is evoked. One might imagine oneself huddled in a winter coat against the bitter winter winds, with one's face going numb from the cold. Or one might imagine oneself working alongside future colleagues whom one has already met at interview. The overall sequence of reflection will result from a complex amalgam of mentally rehearsed actions together with images called into working memory from searches of long-term memory using targeted attention. In any case, here, too, the process is thoroughly action-like, with the overall sequence intentionally controlled by the goal of reaching a decision.

When one engages in prospective forms of reflection, the images that one entertains in working memory will be made available to affective and evaluative networks, among others, as a result of being globally broadcast. These systems respond with some degree of positive or negative affect. This might include some of the bodily changes constitutive of the affective states in question (some visceral, some motor), or swift "as-if" predictions thereof, which can be monitored and used as cues for decision making (Damasio, 1994). But much more important, in my view, is the valence component of affect, which will lead the represented scenarios to seem good or bad to one as one reflects (Carruthers, 2011a). The result is that one option will come to strike one as intuitively better than the other, leading one to choose it.

Orbitofrontal cortex is known to be the main terminus of affective processing in the brain, and damage to this region is known to result in greatly impoverished decision making (Damasio, 1994; Rolls, 1999). Strikingly, patients with orbitofrontal damage can still reason perfectly sensibly about the options open to them, and can offer well-articulated reasons for and against each. Indeed, in many cases their theoretical judgments of which option would be better are quite normal. But their actual decisions fail to reflect those judgments. It seems that there is an

immense difference between explicitly *judging* that an option is good and *seeing it as such*. The latter is constituted by the valence component of affective processing, and seems to depend especially on the normal functioning of orbitofrontal cortex. It appears that valence is needed to provide the motivation for making one decision rather than another.

We also know that when explicit forms of decision making, involving articulated reasons, are pitted against the intuitive affective results of prospection, the latter will often prove superior. This is especially likely where the good-making features of a choice are not obvious, or not easy to articulate (Wilson et al., 1993), or when multiple good-making features need to be combined to reach an overall evaluation (Dijksterhuis et al., 2006). It seems that one of the adaptive (but also potentially misleading) features of the affective system is that valence produced by multiple properties of a situation or thing can be summed together into an overall affective intuition, in ways that are hard to mimic through discursive reasoning (Carruthers, 2011a). In such cases one will just *see* one option as better than another while being unable to articulate why.[5]

Another adaptive feature of affect-based prospective decision making can be inferred from the mechanisms governing entry into working memory, discussed briefly in Section 2. We noted that searches of memory (and probably also mental rehearsals of action) issue in bottom-up forms of relevance-competition. Multiple representations will generally become active in response to any specific memory cue (and multiple action schemata will likewise become active as one searches for actions that might realize a goal). These will compete with one another through the bottom-up attentional mechanism to attract the spotlight of top-down attention and thus enter working memory. As we noted earlier, these competitive processes will involve tacit judgments of value and of relevance to current goals. The images that enter working memory when one engages reflectively in prospection will thus already have been pre-selected as especially relevant to the task or decision in hand.

The upshot of our discussion in this section is that the sensory-based model of reflection has the resources to explain many familiar facts about the nature of reflection, together with a number of less familiar ones as well. Moreover, all of the processes we have considered turn out to be action-like in nature, resulting from intentionally controlled forms of mental rehearsal and attention-allocation.

[5] Or rather, one's attempts at articulation will often fail to track the actual adaptive value of the decision. One can always confabulate an answer, of course, and people in such circumstances often do.

4. Generating Inner Speech

Much of the time, when we reflect, our reflections involve inner speech.[6] The active character of inner speech has already been discussed in Chapter 4.4. We noted that it depends on the mental rehearsal of speech actions, issuing in forward models of the likely sensory consequences of those actions (normally, heard speech). When attended, these sensory predictions are globally broadcast and processed by the language comprehension system. The result is that one *hears oneself as* wondering or asserting some specific content in inner speech. It seems plain, then, that this form of reflection is not only active in nature but sensory based.

On the account of reflective thinking being defended in this book, the contents of working memory (especially when stimulus-independent, as is inner speech) are both controlled and motivated by one's goals and values in the light of current or recently salient information. To be plausible, this should come paired with a mirroring account of the production of overt speech. For it seems quite unlikely that inner speech should be controlled by unconscious forms of decision making whereas overt speech is not. One aim of the present section is to determine whether this commitment to parallelism between inner and outer speech is plausible. Another is to provide a sketch of how covert uses of speech can constitute familiar forms of reflection.

4.1. Speech as a form of unconscious goal pursuit

Standard models of speech production take their start from a communicative intention, or a *thought to be expressed* (Levelt, 1989). But of course speech is generally in the service of wider goals. One may be attempting to persuade someone of something, attempting to establish a rapport with a potential business collaborator, or attempting to impress a potential employer. So at the very least, one's communicative intention is always to express a given thought in the service of some further goal or goals. Sometimes these goals are ones that the speaker knows about, and could articulate if asked.[7] But often they are not. The literature in social psychology is rife with effects on people's speech behavior of goals that speakers are presumably unaware of at the time. These include self-presentation effects, desires to fulfill (or frustrate) the goals that they tacitly

[6] Note, however, that proportions of inner speech, visual imagery, and other forms of stimulus-independent thought vary significantly between people (Heavey & Hurlburt, 2008).

[7] Let me stress again that this does not mean that those goals are conscious. For consciousness requires global broadcasting, rather than knowledge grounded in inferences from ancillary cues.

attribute to their interlocutor, and so on (Kunda, 1999; Moskowitz, 2005; Fiske & Taylor, 2008). For example, people who expect to have their opinions on a topic challenged will modify the views that they express in the direction of those of their audience, presumably with the goal of avoiding confrontation (Cialdini & Petty, 1981).

For a clear example of the effects of unconscious (and competing) goals on speech behavior, we can return to the well-validated findings of the counter-attitudinal essay paradigm, discussed in Chapter 5.2. Recall that after writing an essay arguing for the contrary of what they really believe people will (if their freedom of choice in writing the essay is made salient to them) thereafter express a belief that is much closer to the view they have been defending. We know that they do this in order to make themselves feel better about what they have done, or to present their actions in a better light. Yet it is quite unlikely that this goal is a conscious one. For if it were, and people were aware of saying something other than they believe, then one would expect them to feel *worse* as a result, not better. But notice that the goal of saying what is true (or saying what one believes) does not cease to have any influence at all. If it did, then one would expect people to report the *opposite* of what they believe (that is, the same as the belief they have been arguing for). For that would present their behavior in the *best* light. Rather, people's expressed beliefs are generally somewhere around the middle of the scale between what they really believe and the view they have just been defending. This is best understood as resulting from an unconscious *compromise* between two incompatible goals.

Items of information, too, can compete for expression outside of one's awareness. And sometimes distinct and incompatible items of information can be expressed simultaneously (in different modalities) in the service of the very same goal. This is nicely illustrated by the work of Goldin-Meadow (2005) on gesture. Children who are trying to explain how to solve a math problem will sometimes *say* that the numbers should be combined in a certain order (incorrectly) while at the same time their hand gestures reveal knowledge of the correct groupings. This is especially likely to happen among children who are on the cusp of being able to answer correctly. It seems that the goal of saying how the numbers should be grouped can lead one to express incompatible beliefs in different modalities simultaneously.

Speech is, of course, a form of action. And like other forms of action it can be undertaken in the service of multiple goals influenced by multiple forms of information, often outside of the awareness of the speaker. These points can nevertheless be rendered consistent with the standard model (Levelt, 1989), provided that all of this decision making takes place prior to the formation of a

communicative intention (that is, prior to selection of a thought to be expressed). But it seems much more likely that speech production (like speech comprehension; Hickok & Poeppel, 2007) proceeds in parallel (or at least interactively; Nozari et al., 2011), with decisions about *what* to say being taken while one is in the process of saying it (Dennett, 1991; Lind et al., 2014).

This suggestion is consistent with the findings reported by Novick et al. (2010), that patients with damage to Broca's area (leading to a form of production aphasia) also show much wider deficits, especially in their capacity to inhibit prepotent actions. (For example, they perform quite poorly in the Stroop test.) For such people's expressive difficulties emerge only in cases where there are many competing things they could say. For example, when asked to generate verbs associated with a given noun, patients with damage to Broca's area may become paralyzed when prompted with "ball," since there are many related verbs to choose from ("throw," "kick," "pass," "catch," and so on), while performing as normal when prompted with "scissors," which is associated with just a single action ("cut"). Similarly, healthy people given the same test show increased activity in Broca's area when selecting a verb out of many alternatives, as well as during conflicting-action trials of the Stroop test. At the very least these findings establish that speech production involves competition among expressive *actions*, as well as competition between thoughts to be expressed.

The competitive processes involved in speech production are generally unconscious, of course. Indeed, just as value is processed unconsciously during bottom-up perceptual processing, often issuing in decisions to shift attention from one thing to another, so too is it processed in competition among motor schemata. Decisions about what to say, what words to use, and so on are generally taken "on the fly," resulting from swift evaluations of the competing alternatives. The same is then presumably true in connection with inner speech, thus meshing quite nicely with the account of reflective thinking being presented here.

4.2. Goals in inner speech

Sometimes the goals that guide and motivate inner speech are precisely those that would guide and motivate overt speech (except that something must motivate suppression of overt speech actions in the former case). For sometimes inner speech takes the form of an imagined conversation, or at least speech directed toward an imagined audience. Sometimes one rehearses (and thereby consciously evaluates) what one could say, or should say, or could *have* said to someone. These cases fit the mold of mental rehearsals of action generally, especially as used in prospection. And the same goals will then be in play as would figure in overt actions of the same sort. It seems unlikely, however, that inner speech is

always of this kind. In particular, what goals motivate one's production of inner speech in cases where there is no imagined audience?

Sometimes inner speech is directed toward the solution of some problem. This is supported by the findings of so-called "think aloud" protocols, undertaken while people attempt to solve problems that admit of systematic (uncreative) solutions. (In cases requiring a creative solution, thinking aloud seems actually to impair performance; Schooler et al., 1993.) In these circumstances speech productions turn out to mirror one or another of the objectively determined strategies toward a solution, and the time-course of the sequence maps smoothly onto what happens when people are asked to solve the same problem silently (Ericsson & Simon, 1993). It seems that in such cases one's goal, in producing any given sentence, is to state the next step toward a solution. This issues in a speech act that is globally broadcast, interpreted, and evaluated for relevance and likely success. So there need be no communicative intent, and no goal of producing an effect in an audience. Here the function of inner speech seems comparable to the function of writing that is undertaken for one's own benefit, as when one writes solely to work out one's thoughts. For by providing a sentence that can be consciously perceived, one recruits all of the brain's circuits to evaluate it, and to compete with one another in offering suggestions about what should come next.[8]

It seems that much inner speech meshes nicely with our model of the active, sensory-based, nature of reflective thinking. But some inner speech is not so obviously purposeful. Sometimes, especially during episodes of mind wandering, seemingly random sentences and phrases appear suddenly in consciousness, apparently out of nowhere. Such cases will be discussed in Section 5. But it is likely, I will argue, that they reflect unconscious evaluations of the relevance of competing speech motor schemata under conditions of low cognitive control. The result is a seemingly uncontrolled sequence of speech. But in reality, each item is motivated by a momentary unconscious decision.

4.3. Pragmatics in inner speech?

If inner speech is like outer speech, only with the overt motor component suppressed, then we should expect the meaning of utterances in inner speech to have a rich pragmatic component, just as does outer speech. This is almost certainly true of inner speech that occurs in the context of imagined conversations (although I know of no systematic studies of the question). Based on my

[8] Whether people *intend* these effects when they engage in inner speech is another question, however. Some considerations relevant to this question will be presented in Section 5.

own introspective evidence, at any rate, not only does one seem to take account of earlier stages of the imagined conversation (when selecting pronouns and so forth), but one's inner speech can be rich in irony, sarcasm, and other prototypically pragmatic communicative devices.

In the case of purposeful, problem-directed, speech, too, one would expect to see pragmatic effects. For the processes that generate such speech should be sensitive to the same contextual factors (the perceived layout of the environment, memories of recent utterances, and so on) that operate when one speaks to another person. Admittedly, when there is no audience (real or imagined) there is no need to take into consideration the knowledge or ignorance of the other person, and their differing perspective on the situation. Nor is there any need to build a representation of "common ground" that one is said to draw on when producing utterances for an audience (Stalnaker, 2002). But such representations seem to play a much smaller role in language production than was once thought. For speakers often default to their *own* perspective on the situation when selecting utterances, needing to inhibit this tendency if they are to take account of others' differing perspectives (Horton & Keysar, 1996; Shintel & Keysar, 2009).

It is hard, of course, to undertake direct experimental studies of the role of pragmatics in inner speech. But on the assumption that people engaged in think-aloud protocols make utterances that faithfully reflect what they *would* have said silently to themselves otherwise, one can study transcripts of what people say aloud in such circumstances. And in fact such transcripts are rife with pragmatic effects. Indeed, without knowledge of the exact circumstances, one can frequently not understand what the speaker is referring to at all. Consider, for example, just one episode selected from among the raw data of a think-aloud protocol used while participants solved a set of Raven's Matrices (kindly made available to me by Mark Fox).[9] One person said this: "Oh, Um, for each there's a diamond. Um, a diamond. And one going the other way."

Reading this, one has to ask: for each *what*? And one *what* going the other way? Looking at the matrix that this person was solving (see Figure 6), one can figure out that he means each *line* in the matrix contains a diamond. And what "goes the other way" are a set of diagonal lines that bisect each figure on each line, one upright, one slanted to the left, and one slanted to the right. Indeed, when thus interpreted, the two statements strongly suggest a particular solution to the problem. Since he did, indeed, get the answer right, it seems likely that the

[9] Raven's Matrices are often used in intelligence tests. They comprise sequences of geometric figures, and the participants' task is to select the next shape in the sequence from a number of alternatives. See Figure 6.

person's speech played a role in enabling him to solve it. Indeed, the episode strongly suggests that one of the functions of inner speech is to fix in working memory some of the components of the solution to a problem while one continues to consider others.

If speakers lack introspective access to their communicative intentions while speaking, as I have argued (Carruthers, 2011a), then how do participants who are thinking aloud (or entertaining the same sentences in inner speech) know how to interpret their own utterances? Many of the same contextual cues are available in the first-person as in the third, of course. And interpretation can also be guided by knowledge of one's own focus of attention. In the episode described above, for example, the person was presumably attending to the rows while saying, "For each there's a diamond," and was attending to the diagonal lines in the figures while saying, "And one going the other way." We, too, would have known how to interpret these utterances had we been aware of what he was attending to at the time.

Overall it appears that the sensory-based model of reflection has the resources to explain some of the familiar properties of inner speech. And the close parallels between inner and outer speech support the view that the former is active in nature, and results from goals and values that do not themselves figure among the contents of working memory.

5. Creativity and Mind Wandering

The topic of the present section is uses of working memory that do not seem, introspectively, to be active in nature. Sometimes one's thoughts change direction for no apparent reason (especially when one's mind is wandering). Sometimes ideas seem to leap into mind unbidden (particularly when those ideas are novel or creative). It seems to us in such cases that we are passive receivers of our own thoughts, rather than agents who actively produce or control them. Indeed, it seems to us that the stream of consciousness when we mind wander flows according to some unknown set of forces and influences, rather than being actively produced. I will argue that this impression is a mistake. I will also consider what can be said about the nature of creativity, from this perspective, as well as what can be said about why we engage in mind wandering at all.

5.1. The active nature of mind wandering

There are strong negative correlations between working-memory capacity and the extent to which people mind-wander during attentionally demanding tasks. This has been shown using introspection-sampling probes to establish the

presence of mind wandering, both during experimental tasks of varying demandingness (Kane & Engle, 2003) and during everyday life (Kane et al., 2007). Moreover, the capacity to sustain attention to a task seems to involve two distinct components. One is the ability to maintain a goal in an activated state over stretches of time when it does not need to be acted upon (such as watching out for an "oddball" stimulus in long sequences of predictable stimuli), and the other is the capacity to direct attention in the light of that goal while resisting interference from other factors (McVay & Kane, 2009).

On the face of it, such findings might seem to undermine, rather than to support, the claim that mind wandering is active in nature. For mind wandering seems to occur when controlled attention to a task fails. In fact, however, there is continual competition between the "task positive" (attention-controlled) and the "task negative" (default) networks (Kelly et al., 2008) mediated by the bottom-up attentional system (Yeo et al., 2011). While the latter is probably *always* active, monitoring both peripheral stimuli and partially activated endogenous representations for relevance to one's values and current goals, its influence is especially noticeable during mind wandering. I suggest that during such episodes the bottom-up system is winning the competition for top-down attentional resources, wresting control of that system away from the current task-dependent goal (if there is such a goal).

Consistent with this suggestion, the region of medial prefrontal cortex that constitutes part of the default network is anterior cingulate (Buckner et al., 2008), which is known to play a crucial role in conflict monitoring (Botvinick et al., 2004; Kerns et al., 2004). Indeed, the most recent model of the function of anterior cingulate is that it makes decisions about the allocation of top-down control based on cost–benefit analyses of the conflicting alternatives (Shenhav et al., 2013). Also consistent with the proposed active nature of mind wandering, the executive components of top-down attentional systems are also active during mind wandering (Christoff et al., 2009; Smallwood et al., 2011; Christoff, 2012), especially those located in dorsolateral prefrontal cortex.

We can hypothesize, then, that during mind wandering there is no single goal controlling the direction of top-down attention in a sustained way. Rather, contents that have been partly activated through associative and other uncontrolled processes are evaluated by the bottom-up attentional system, competing with one another (and with the task goal, if there is one) for control of the top-down attentional system and subsequent entry into the global workspace. Sometimes one of these contents wins the competition and is judged most relevant to one's values or goals, issuing in a decision to redirect top-down attention accordingly. But in the absence of a strong enough sustained goal, that

conscious content or series of contents is likely soon to be supplanted by another. As a result, one's thoughts when mind wandering may flit from topic to topic. But each individual content or short sequence of contents nevertheless results from a momentary decision to redirect top-down attention in the light of a judgment of relevance. Each therefore results from a process that is action-like in nature. So while the entire sequence of contents during mind wandering is not goal directed or actively controlled, each individual content is.

Mind wandering is active, I suggest, in much the same sense that someone *physically* wandering around in a garden is active. Such a person's movements are actions governed by momentary decisions—now to walk here, now to walk there—even though there is no overarching goal that governs his actions (beyond, perhaps, the goal of allowing himself to wander at will).

Moreover, recall that mind wandering often involves inner speech. Since the latter results from mental rehearsal of speech actions, it, too, is active in nature, even when not controlled by a sustained overarching goal. It seems likely that just as sensory-based representations compete with one another through the bottom-up attentional network to gain entry into working memory, so is there also competition among potential speech actions for rehearsal and subsequent global broadcast. Current or recent contents of working memory will prime one for speech, and in people who habitually engage in inner speech there will be bottom-up competition among the speech acts one could activate in response. The one deemed most relevant will be rehearsed, and its forward model, when attended, will be consciously experienced. So although it can seem to one during mind wandering that one simply *finds oneself* entertaining sentences in inner speech, in reality the latter are actively caused and controlled.

When our minds wander, we often have the impression that we are passive with respect to our own conscious thought processes. (This is especially likely to be true when our minds wander in ways that strike us as novel, creative, or unexpected, as we will discuss shortly.) It is not *us* who controls the contents of our minds in such circumstances, we think, but some set of forces outside of us. But this impression is an illusion. In these cases, too, decisions are taken about the contents that are most relevant to our goals or concerns, and those decisions control the direction of top-down attention. The difference between cases where our thoughts seem to us to be actively controlled and those where they do not is just the difference between cases where we think we know immediately *why* we entertain the thoughts that we do, and those where we take ourselves to be ignorant of those reasons.

5.2. *Explaining creativity*

Creative episodes can be divided into two main components or phases, as emphasized by so-called "GENEPLORE" (for "generate and explore") models of creative cognition (Finke et al. 1992; Finke 1995; Ward et al. 1999). The first is the generation of candidate ideas, and the second is the evaluation and development of those ideas before acceptance or implementation. (In practice, of course, the two phases will be intermixed, as one generates and evaluates one idea while also considering others.) As one might expect, the evaluative, analytical, phase of creativity is associated especially with activity in executive and top-down attentional regions of cortex, including dorsolateral prefrontal cortex (Ellamil et al., 2012). Our focus here will be on the generative component of creativity, since it is this that gives rise to the feeling that we are passive with respect to the emergence of new ideas.

Numerous forms of data suggest that the generative phase of creativity results from decreased or defocussed top-down attentional control. Thus (and consistent with folk beliefs) we know that numerous factors besides the intention to be creative (Baumeister et al., 2007) can influence levels of creativity. Alcohol increases creativity (Jarosz et al., 2012); sleepiness increases creativity (Wieth & Zacks, 2011); and being in a good or relaxed mood improves creativity (Subramaniam et al., 2009). Direct-current stimulation applied to left prefrontal cortex also increases creativity (Chrysikou et al., 2013), presumably by down-regulating top-down attentional control.[10] And musicians engaged in creative improvisation show lesser activity in dorsolateral prefrontal cortex than do those who are playing a tune they know well (Limb & Braun, 2008). Moreover, particularly creative individuals tend to be lower in what is called "latent inhibition," which is the capacity to screen from awareness stimuli previously experienced as irrelevant to one's task (Carson et al., 2003).

Such data might seem to support a passivity view of creativity, since they show that creativity happens when top-down attentional control is relaxed. But, as previously, the data are equally consistent with an account in terms of bottom-up forms of attention competing to influence the direction of top-down attentional networks in the absence of a strong and active goal, with the top-down network being controlled by momentary decisions of maximal relevance. For example, here is what might happen when someone searches for unusual things one could do with a brick. The intention to be creative leads to a defocussing of attention,

[10] Direct-current stimulation, like transcranial magnetic stimulation (TMS), suppresses the level of neural activity in regions to which it is applied.

removing the suppressing effect on remotely associated representations that are activated (albeit only weakly) by the thought of a brick. (On the contrary, it is the obvious and familiar uses of a brick—such as building a wall—that need to be suppressed.) These associated representations are received by the region of ventral parietal cortex that forms part of the bottom-up attentional network. This communicates back and forth with right ventrolateral prefrontal cortex to evaluate their degree of interest. The result is a competition to gain control of the top-down network. This takes place in the medial frontal gyrus and anterior cingulate cortex, issuing in a decision to redirect the top-down network toward one set of representations in particular. The result is that a particular idea pops into the person's conscious mind, seemingly unbidden, thereby enabling a reflective, conscious, evaluation of its merits.

Is there evidence to support such an interpretation? There is some. One imaging study of creativity, for example, found a trade-off between frontal and occipital–temporal cortical activity depending on whether the task was mundane (greater frontal) or creative, suggesting lesser top-down control in the latter case (Chrysikou & Thompson-Schill, 2011). Another found increased activity in anterior cingulate cortex in creative conditions (associated with conflict monitoring and decisions about the direction of top-down control), as well as in the medial central gyrus of right prefrontal cortex, which is a component of the bottom-up attentional network, as we noted in Chapter 3.4 (Howard-Jones et al., 2005). Moreover, numerous studies of creativity have found associated activity in ventral parietal cortex, which is another crucial component of the bottom-up attentional network (Bekhtereva et al., 2004; Jung-Beeman et al., 2004; Geake & Hansen, 2005; Subramaniam et al., 2009).

While more evidence would be welcome, it seems that current findings are at least consistent with the active nature of creative idea-generation sketched here. Moreover, since the level of activation of unattended but weakly active neural representations will fluctuate stochastically, the present account can accommodate quite naturally the partly stochastic nature of creative idea-generation (Simonton, 2003). For as one rather than another representation becomes more active, so will it have a greater impact on the competition for control of top-down attention.

In previous work I urged the benefits of a theory of creativity that is action-based (Carruthers, 2007, 2011b). Among other things, I argued that known properties of the motor system can help explain the partly stochastic nature of creativity. For at some levels of representation action selection is itself stochastic in nature (Rosenbaum et al., 2001, 2008). I also argued that an action-based account enables human creativity to be situated phylogenetically, as an

exaptation of the kinds of protean motor behavior exhibited by many species of animal (especially when escaping from a predator; Driver & Humphries, 1988; Miller, 1997). These properties, I suggested, are inherited by human speech-production systems, issuing in the striking creativity of much overt (and inner) speech. But these suggestions were made before I had come to realize that attention is itself a form of action. The present account can thus inherit all of the advantages of my previous one, but broadened to encompass creative imagery quite generally.

The main point of the present discussion, however, is just that one can provide an active, sensory-based, account of creative idea generation that is consistent with (and to some degree supported by) the existing evidence.

5.3. Why mind wander?

Why do people mind wander at all? The question can be understood either distally or proximally. Consider the former first. From a functional and evolutionary perspective it might seem puzzling that people's minds should often drift from the here-and-now, meandering across past, future, and merely possible scenarios instead. But only a little reflection is needed to dispel any air of mystery. There are frequently lessons to be acquired from the past that were not learned at the time. Recalling those episodes, thereby making them globally available in working memory, can issue in novel inferences and affective reactions. The results can in turn be stored for future use. Moreover, the value of prospective reasoning has already been emphasized. By entertaining representations of various future actions in working memory one can predict their likely consequences and respond affectively to them, thereby enabling one to plan (Damasio, 1994; Gilbert & Wilson, 2007). Even merely hypothetical or counterfactual scenarios can give rise to potentially useful information. So the adaptive nature of mind wandering (at least in the absence of any significant current task) should not be in doubt. In Chapter 8 we will consider whether, and to what extent, mind wandering has evolutionary precursors.

Now consider the proximal causes that initiate episodes of mind wandering. In light of the above, it makes sense that there should be a strong disposition to begin mind wandering as soon as one finds oneself without any current task, such as lying in an fMRI scanner while waiting for instructions from an experimenter or (more mundanely) when waiting for a bus. For by having such a disposition, any "down time" of this sort can be put to adaptive use. What in turn grounds this disposition might be that people find mind wandering inherently rewarding (Picciuto, 2011), in addition to any content-related rewards they might experience. (Of course the contents one entertains when mind wandering are often—

but by no means always—experienced as pleasurable. Think, here, especially of fantasy.) It is more puzzling, however, why one should so often mind wander when engaged in an important task. Why is it often so hard to keep one's attention on one's current task, even when one is highly motivated to do so? And what causes one to shift attention away from that task when one does?

Kurzban et al. (2013) provide a general model in terms of which this issue can be understood. They argue that the experience of intellectual effort is an affective response designed to make one's current activity seem bad, grounded in a cost–benefit analysis of that activity in comparison with what else one could be devoting one's attention to at the time (including mind wandering). They amass an impressive range of evidence in support of their model, while contrasting it favorably with competitors. In effect, the idea is that when one switches attention away from one's current task and initiates an episode of mind wandering, one has taken a *decision* that the latter activity is the better of the two in the circumstances. This decision is not conscious, of course; and people are generally unaware that they have made it, or why. This is why one seems to just *find oneself* mind wandering, passively, often issuing in conscious regret. But in fact, initiating mind wandering is an action (or is at least action-like), just as are the individual movements of attention that determine the latter's contents. Further confirming the correctness of this account, a very similar model is presented independently by Shenhav et al. (2013). They propose that the computational function of anterior cingulate cortex is to take decisions about the allocation of top-down control mechanisms on the basis of cost–benefit analyses of the various options.

In conclusion, then, a case can be made for saying that all aspects of the stream of consciousness are really active in nature. Not only is this true, most obviously, of inner speech and the mental rehearsal of action generally (for example, in prospection), but it is also true of forms of reflection that seem intuitively to be passive. Even the genesis of creative ideas and seemingly undirected mind wandering result from decisions to redirect top-down attention in the light of cost–benefit judgments or judgments of relevance.

6. Conclusion

This chapter has outlined a sensory-based, working-memory-based, account of conscious reflection and the stream of consciousness. On this view one's amodal attitudes operate entirely in the background, helping to determine the contents of working memory and the direction that one's reflections will take. Moreover, the stream of consciousness is thoroughly active in nature, at least to the extent that its momentary contents are under intentional control. This is because it always

depends on the direction of attention, and often depends on mental rehearsals of action, both of which are action-like in nature.

Our common-sense belief that reflection is often an active process is thus vindicated. But common sense also maintains that conscious thinking is often *not* active, because thoughts and images enter our minds and interrupt to our reflections unbidden, seemingly appearing out of nowhere. According to the view outlined here, this aspect of common-sense belief is mistaken. Ideas that appear seemingly from nowhere are really a result of unconscious decisions that resolve conflicts over the direction of attention, or concerning what actions to rehearse. The difference between the cases that common sense regards as being under our control and those that it maintains are not, is simply that in the former cases we think we know *why* we reflect as we do, whereas in the latter cases we don't. This difference is merely epistemic, however, and demonstrates nothing about the real nature of the causal processes that produce the conscious contents in question.

Recall from Chapter 2.3 that philosophers who claim that our amodal attitudes are sometimes active and under intentional control are committed to the view that such attitudes admit of two distinct varieties: a set of action-like attitudes, and a set of passive ones invoked to explain the active status of the former set. As we saw in Chapter 2.3, this commitment gives rise to a number of problems. Work needs to be done to explain why active and passive instances of belief, for instance, should nevertheless be considered as belonging to the same mental-attitude kind; for the functional roles of those two sorts of instance will be quite different. And some sort of complex mental architecture will need to be postulated to explain how some propositional attitudes can participate in the intentional control of others. Since we have no knowledge of such an architecture, this aspect of the stream of consciousness would remain currently unexplained.

The sensory-based model of reflection makes no such commitment, of course, and gives rise to no such problems. It holds that amodal attitudes are never under direct intentional control, whereas the stream of consciousness always is; and the latter is explained in terms of mechanisms that are already known to exist, and about which much is already known. The sensory-based model therefore has greater explanatory adequacy, as well as an attractive simplicity and elegance that the philosophers' account lacks. This amounts to yet another strike against the latter. In addition to all of the direct evidence against the idea of an amodal central workspace and in support of its sensory-based competitor, reviewed over the course of previous chapters, the sensory-based model of reflection is also favored by considerations of simplicity and explanatory adequacy.