

*Inner Speech and Outer Thought**

Keith Frankish

1. Introduction

Recent years have seen extensive empirical studies of many aspects of inner speech, including its phenomenology, cognitive effects, neural basis, and pathology (for a survey, see Alderson-Day & Fernyhough, 2015). Despite this, however, general questions about the nature and function of inner speech remain unsettled, and there is no universally agreed theoretical framework for empirical work. Here, Martínez-Manrique and Vicente have distinguished two contrasting theoretical approaches — the *format view* and the *activity view* — which provide useful reference points in locating views on the matter (Martínez-Manrique & Vicente, 2015). According to the format view, inner speech is an output of the language production system, and its function is to enable conscious thought by supplying an appropriate format, or vehicle, for it. According to the activity view, inner speech is an activity (‘innerly speaking’), which is not necessary for conscious thought and has many functions, continuous with those of outer speech. These views embody contrasting accounts of the cognitive role of inner speech. The format view holds that inner speech *is* thinking; episodes of inner speech are episodes of reasoning, judging, or deciding. The activity view, by contrast, holds that inner speech *assists* thinking; episodes of inner speech are episodes of self-prompting, self-reminding, and self-commentating, which may aid reasoning, judgement, and decision making.

So defined, the two views are incompatible (one claims that inner speech is necessary for conscious thought and the other claims that it is not), and Martínez-Manrique and Vicente focus on arguing for the activity view over the format view. I think this is too hasty, however, and I shall argue that we can — and should — endorse modified versions of both views. I’ll define these modified versions shortly. First, however, I want to acknowledge a challenge to any attempt to reconcile the two perspectives. If inner speech consists of inner utterances, as the activity view holds, then it is hard to see how it can constitute a form of thinking, as the format view has it. Utterances are typically intentional actions — actions performed for a reason. (For present purposes, I shall take it that intentional action is action that is guided by (possibly nonconscious) decisions resulting from practical reasoning drawing on one’s goals and beliefs.)¹ But does it make sense to talk of *thinking* being intentional in this sense? What would the motivating beliefs and desires be? Moreover, how could an utterance play the *causal role* of a thought — of a decision, say, or a judgement? A decision produces a settled intention, and a judgement produces a settled belief. How

* This is the author’s preprint of a chapter in Peter Langland-Hassan and Agustin Vicente (eds.), *Inner Speech: New Voices*, Oxford University Press, 2018. It may differ in minor ways from the print version, which is the definitive one. Version of 07/12/17.

¹ For defence of the claim that speech is typically goal-directed, see Carruthers, 2015, pp. 161–3.

could an utterance have such effects? For example, how could saying to myself, ‘Interest rates will rise soon’ produce in me the belief that interest rates will rise soon — a persisting mental state that plays a role in my subsequent reasoning and planning? We usually think of utterances as expressions of our mental states, not causes of them. Doesn’t the view I am proposing get the causal sequence the wrong way round?

I think there is a good response to this challenge, which I shall set out in the course of this chapter. It will involve drawing on contemporary dual-process theories of reasoning to develop a view of conscious thinking as an intentional activity, which can be conducted overtly, using outer speech. When inner speech has a cognitive function, I shall argue, it is simply an internalization of such ‘outer thinking’.

The rest of the chapter is structured as follows. Sections 2 and 3 are devoted to identifying the versions of the format and activity views to be defended and showing that they are formally compatible. Section 4 discusses how inner speech might enable conscious thought, rejecting the view that it merely makes thoughts conscious and arguing that speech-dependent thinking must be a distinct *type* of thinking. Section 5 introduces dual-process theory, according to which humans engage in two different types of reasoning, and section 6 argues that the core difference between the two types is that one involves intentional activity. Section 7 discusses how outer speech can implement active reasoning processes, proposing a cyclical model in which utterances serve to break down complex problems into manageable subproblems, and section 8 shows how judgements and decisions can also be linguistic acts.

2. Inner speech as format

As Martínez-Manrique and Vicente define it, the format view comprises three claims (‘IS’ means ‘inner speech’):

- (i) the strong consciousness thesis: IS is necessary for conscious thinking;
- (ii) the format thesis: in IS we recruit a representational system because of its features as a format;
- (ii) the product thesis: IS consists in some output of the linguistic production system, typically strings of phonological representations.

(Martínez-Manrique & Vicente, 2015, p. 1)

Martínez-Manrique and Vicente note that (i) and (ii) are counterparts: the first says that conscious thinking requires a certain format, or vehicle, and the second that this format is supplied by inner speech. The format view does not specify what kind of vehicle is required, and it is compatible with a range of views on the matter. The third component of the view, the product thesis, is to be understood as contrasting with the claim that inner speech is an activity (Martínez-Manrique & Vicente, 2015, p. 2). The idea, I take it, is that inner speech, unlike outer speech, is not under intentional control. Martínez-Manrique and Vicente suggest that this thesis is a consequence of the previous two, which they regard as the core of the format view (2015, p. 2). Finally, Martínez-Manrique and Vicente take the format view to involve the claim that the

proper function of inner speech is to enable conscious thinking and that any other functions it may have are derivative from this one (2015, p. 6). I'll call this *the proper function thesis*.

There are reasons to endorse the format view. Conscious mental states, I assume, are known to their possessor in a direct way; we do not have to infer their existence from sensory evidence, as we infer the existence of other people's mental states. (There are different accounts of what it is for a propositional attitude to be conscious, but on most views of the matter this assumption will hold; see Carruthers, 2011, pp. 373–9.) But there is a strong case for thinking that the only mental states to which we have such direct, non-inferential access are ones that are sensory. For example, Peter Carruthers has argued that we have no special faculty of self-knowledge and that our knowledge of our own mental states is dependent on the same mental system that gives us knowledge of other peoples' minds — a system that evolved for the purposes of social cognition and has access only to sensory inputs (Carruthers, 2011). It is a consequence of this view — which Carruthers calls the Interpretive Sensory-Access (ISA) theory of self-knowledge — that we have direct access only to mental states that are sensory, or sensorily embedded, and that knowledge of our non-sensory mental states is derived by inference from sensory information (broadly construed to include sensory imagery as well as perception). Given the assumption that conscious mental states are ones to which we *do* have direct, non-inferential access, it follows that conscious thoughts, if they exist at all, must have a sensory vehicle, or format, of some kind — which inner speech might provide. (This view does not assume any specific view of consciousness itself, but for present purposes I shall adopt a 'global broadcast' model, on which sensory states become conscious when targeted by attention and globally broadcast to systems for memory, reasoning, motivation, and decision-making; see, e.g., Baars, 1988.)²

This is not the place to assess the ISA theory. I appeal to it here primarily to motivate serious consideration of the format view. (For related arguments for the view that conscious thought requires a sensory format, see Jackendoff, 1996; Prinz, 2011.) I should stress, however, that Carruthers himself does not adopt the format view. He denies that episodes of inner speech have the right causal roles to qualify as thoughts and argues that (with limited exceptions) none of our thoughts are conscious (Carruthers, 2011, ch. 12). Thus, as well as motivating the format view, ISA theory poses a challenge to it. Since it is my aim to answer this challenge, I shall take the truth of ISA theory as a constraint in what follows.

Although I think the format view is correct in spirit, I want to make some modifications to it. First, the strong consciousness thesis is *too* strong. I do not want to claim that inner speech is strictly necessary for conscious thinking. What *is* necessary, I shall argue, is a symbolic medium that is perceptually or introspectively available to the agent. I think inner speech is the dominant medium of this kind and the source of

² This model may not explain the putative *phenomenal* aspect of consciousness, but, as Carruthers notes, it is widely agreed that global broadcast at least *coincides* with consciousness (Carruthers, 2011, pp. 48–9).

the richness of conscious thinking, but other types of mental imagery could also serve, as well as external symbols, including outer speech.

Second, I reject the product thesis. Inner speech can both supply a vehicle for conscious thinking *and* be an activity. To see how, we can turn again to Carruthers. Inner speech, Carruthers argues, involves the mental rehearsal of actions (Carruthers, 2006, 2011, 2015, this volume). Briefly, the idea is this. When we decide to perform an action, a copy of the efferent motor commands is used to generate a *forward model* of the expected sensory consequences of the action, both somatosensory and perceptual. When the action is performed, this model is compared with the actual incoming sensory data to check that things are going as expected. However, forward models can also be produced without overt action. We can initiate the commands for an action, triggering the construction of a forward model, while suppressing the movements themselves. If the resulting forward model is then targeted by attention, we imaginatively experience ourselves performing the suppressed action and perceiving its immediate consequences. Such mental rehearsal, Carruthers argues, plays an important role in planning, allowing us to try out candidate actions in imagination and assess their effects before deciding whether or not to perform them. Inner speech, Carruthers suggests, results from the mental rehearsal of utterances. When we rehearse an utterance, the forward model generated includes auditory representations of how the utterance would sound. These representations are interpreted by the language comprehension system, and if targeted by attentional mechanisms and globally broadcast, they are experienced as inner speech. The broadcast contents may then have a range of cognitive effects, just like the contents of heard speech.

If this is right, then inner speech can be both format and activity. A mental rehearsal, like the action it rehearses, can be intentional. (Indeed, there is substantial evidence of the value of deliberately rehearsing skilled actions in imagination as practice for actually performing them; see, e.g., Driskell et al. 1994). Utterance rehearsal is no exception; in preparing for a job interview, for example, I may intentionally try out different responses to questions in imagination. And this activity might also generate sensory representations of sounds (an output of the language production system) that serve as a medium, or format, for conscious thought.³ Of course, there is still a contrast between talking of inner speech as an activity and talking of it as a format. The references are to different things — in the first case to an act of mental rehearsal, in the second to something produced *during* the mental rehearsal (attended sensory representations). This is merely an ambiguity in the use of the term, however, akin to the familiar product/process ambiguity.

Finally, I shall reject the proper function thesis. I do not wish to claim that supporting conscious thinking is the sole, or even primary, function of inner speech. I shall say more about this in the next section.

³ Martínez-Manrique and Vicente acknowledge this point and concede that the product thesis is not central to the format view (Martínez-Manrique & Vicente, 2015, p. 2).

These modifications yield a slimmed down, but still substantive, version of the format view, which might be summed up as follows:

Modified format view: One of the functions of inner speech is to provide a format (representational medium) for conscious thinking.

3. Inner speech as activity

The activity view, as Martínez-Manrique and Vicente conceive it, contrasts sharply with the format view. It denies both of the latter's central theses (that inner speech is necessary for conscious thinking and that we recruit it as a format for such thinking), and claims instead that inner speech is an activity, which has many functions, continuous with those of outer speech (Martínez-Manrique and Vicente, 2015, p. 7). (I'll call these the *activity thesis* and the *continuity thesis*, respectively.) Martínez-Manrique and Vicente suggest that this offers a natural way to think of inner speech, and they stress the similarities between inner and outer speech:

There seems to be no deep difference between reasons why we talk to ourselves and reasons why we talk to someone else: we talk to express ourselves, to motivate others, to evaluate events or subjects, to help people to find places, to regulate their behavior, etc. (2015, p. 8)

As Martínez-Manrique and Vicente note, the activity view is similar to that of theorists in the Vygotskian tradition, who see inner speech as a self-directed, internalized version of interpersonal practices of verbal instruction and regulation. Having been accustomed to hearing others guide and assist them with verbal commentary, children start to produce a similar commentary on their own behaviour, first in self-directed outer speech (private speech) and then, in more compressed form, in inner speech, which inherits the dialogic character of the interpersonal commentary it imitates. In this way, inner speech comes to play important roles in self-regulation and planning, among other things. There is good evidence to support this developmental story (e.g., Winsler, Fernyhough, & Montero, 2009; Winsler & Naglieri, 2003).

Again, I agree with the spirit of the activity view. We have already seen that it is plausible to think of inner speech as involving acts of intentional mental rehearsal. And if inner speech is an activity, then it is also plausible to think that the uses to which we put it (the functions it has) are continuous with the uses to which we put outer speech — advising, approving, asserting, directing, encouraging, informing, persuading, promising, ordering, questioning, reciting, reminding, and so on. Anyone who claims that we use inner speech for purposes that have no continuity with the uses of outer speech faces a difficult question: How did we learn to use it in those ways? It's hard to see how it could have been through instruction or imitation. There is no mystery, however, if our habits of inner speech are modified versions of our habits of outer

speech, which we learned and practised in social contexts.⁴ This isn't a conclusive argument, of course, and it is compatible with speech habits being transformed during the internalization process, both in form and to some extent in function, but it suggests that the burden of proof lies with anyone who thinks that inner speech *does* have wholly new functions.⁵

So, I am in broad agreement with the activity view. As with the format view, however, I also want to make some modifications to it. First, although I think that most episodes of inner speech are intentional actions and shall be concerned only with those that are, I shall not insist that all are. Perhaps some are involuntary, like a reflex. (Certainly, there is a case for thinking that not all are *experienced* as active; it has been argued that auditory verbal hallucinations are misattributed episodes of inner speech, which are experienced as coming from an external source; e.g., Bentall, 1990; Frith, 1992.) Second, I shall drop the negative component of the activity view — the denial of the strong consciousness and format theses. My aim is precisely to show that the positive component of the activity view is compatible with those theses — or rather with the modified versions of them incorporated in the modified format view. This gives us the following modified version of the activity view:

Modified activity view: Inner speech is typically an activity, which has many functions, continuous with those of outer speech.

4. Thinking as self-communication?

Having sorted out the definition of the two views, I turn now to the task of showing how both can be true. We have seen that it is not incoherent to claim both that inner speech is an activity and that it provides a representational format for conscious thinking. But that is the easy bit. It is another matter to show how the format provided by inner speech could actually be *used* for conscious thinking — that is, how acts of innerly speaking could play the role of thoughts. (This is the challenge mentioned in the introduction.) Moreover, we also need to show how inner utterances could play that role if the continuity thesis is true.

To begin with, we need to be clearer about what conscious thinking is and how inner speech might enable us to engage in it. One possibility is that inner speech enables us to

⁴ Deniers of the continuity thesis may reply that the question I pose for them is just a version of a more general question that everyone must address — namely, how did we learn to think consciously? This may be true, but it does nothing to weaken the case for the continuity thesis. For, when combined with the modified format view and developed in the way I shall propose, the continuity thesis offers an answer to precisely that question. (Thanks to Peter Langland-Hassan for raising this point.)

⁵ Peter Langland-Hassan has pointed out that there is also a certain burden of proof on defenders of the continuity thesis. Outer speech usually presupposes a hearer with different beliefs and desires, who needs to be informed, persuaded, and so on, so defenders of the continuity thesis must explain why we should need to speak to ourselves. An answer to this challenge will be developed in the rest of this chapter. In essence, I shall argue that conscious thought constitutes a distinct level of mentality which operates through mechanisms of self-questioning and self-prompting and whose relation to the nonconscious mind has a dialogic character.

bring pre-existing nonconscious thoughts to consciousness: our thoughts are initially nonconscious, but by expressing a thought in inner speech we give it a sensory vehicle that can be targeted by attention and globally broadcast, thereby rendering it conscious.⁶ This process could be intentional; we might express a thought in inner speech (rehearse an utterance that expresses it) because we want to make it conscious — to reflect on it, mull it over, give it more consideration. (This needn't require a sophisticated understanding of the different roles of conscious and nonconscious mental states, just a general sense that we can learn more about our thoughts by making them conscious.) Of course, we are not aware of having desires of this kind, but that is to be expected; the desires that cause our thoughts to become conscious would not themselves be conscious ones. Moreover, this suggestion is compatible with the continuity thesis. In making a thought conscious, we are communicating the thought to ourselves (our *conscious* selves), just as we might communicate it to someone else. (It may be objected that in order to decide to communicate a belief, one must already believe that one has it, and if one already believes that one has it, then one does not need to communicate it to oneself. But there is no reason to think that the act of communicating a belief is guided by a higher-order belief in this way. It is generally agreed that speech production starts with a belief, whose content is the message to be communicated (e.g., Levelt, 1989). One simply has the belief, wants to communicate it, and decides to do so.)

Perhaps some episodes of inner speech do serve as self-communications. However, this is not sufficient to vindicate the modified format view. For self-communications do not have the defining features of conscious thoughts. There are two points to make. First, self-communications do not give us the direct access to our thoughts that is characteristic of consciousness. On the view proposed, it is auditory images of utterances that are conscious, not thoughts themselves, and these images need to be interpreted, just as heard utterances do. As Carruthers stresses, this process may involve, not only decoding the utterance's semantic content, but also theorizing about the intentions behind the utterance — in order, for example, to resolve ambiguities, recover pragmatic aspects of meaning, and work out what attitude is being expressed (Carruthers, 2011, pp. 86–96). Of course, the interpretation process will usually be swift and nonconscious, and there may be many contextual clues available to help it, but (assuming the ISA theory is true) it is always required. Second, acts of self-communication do not play the causal roles of thoughts. Judging that *p* results in my forming the belief that *p*, but self-communicating the thought that *p* does not (I already have the belief). Similarly, employing a thought in reasoning results in my drawing some inference from its content, but self-communicating the thought doesn't. Rather, its main effect is to make me aware that I have the thought — to produce a higher-order belief. Self-communication facilitates self-knowledge rather than first-order reasoning.

⁶ This is how Martínez-Manrique and Vicente understand the format view. As they put it, 'Defenders of the format view hold that by producing strings of phonological representations we bring thought-contents to consciousness' (Martínez-Manrique and Vicente, 2015, p. 4).

This may be a useful cognitive role for inner speech, but it is an ancillary role, not the role of thought itself.⁷

Thus, if inner speech does enable conscious thought, it must do more than act as a channel of self-communication. Rather than just enabling thoughts to become conscious, it must enable a distinctive kind of thought, which is conscious. It must *make conscious thought*, rather than just *making thought conscious*. (In other words, it must provide a vehicle for first-order reasoning —reasoning about things other than the thoughts it putatively expresses.)⁸ And if the activity view is true, this kind of thinking must be an activity, which can be, and initially is, performed overtly, in outer speech. There is, in fact, a strong, independent case for the view that humans exhibit a kind of thinking of just this type. I turn to this now.

5. Thinking as dual

Since the 1970s many psychologists studying reasoning, decision making, and other higher cognitive functions have proposed ‘dual-process’ theories, according to which two qualitatively different processes, or types of processing, compete for control of behaviour (for surveys, see Evans, 2011; Frankish & Evans, 2009; Frankish, 2010; Kahneman, 2011). The theories differ in detail but show considerable overlap. Typically, they distinguish one kind of processing (Type 1), which is fast, automatic, and nonconscious, and another type (Type 2), which is slow, controlled, and conscious. Type 1 processing has also been variously described as low-effort, high-capacity, parallel, contextualized, associative, biased, undemanding of working memory, shaped by biology and personal experience, and independent of cognitive capacity, while Type 2 processing has been described as effortful, low-capacity, serial, decontextualized, rule-governed, normative, demanding of working memory, shaped by culture and tuition, and correlated with individual cognitive capacity.⁹ There are different accounts of the relation between the two forms of processing. A popular view is that Type 1 processing controls behaviour by default, generating rapid intuitive responses, and that Type 2 processing is activated only if circumstances require it (for example, if the subject feels

⁷ Martínez-Manrique and Vicente make the same point:

converting a propositional content into an object one can “look at” only enables subjects to know what they are thinking, not to think those thoughts consciously. Instead of making them aware of a certain propositional content *p*, and so to consciously believe or judge that *p*, this mechanism makes them aware that *they are thinking* that propositional content, i.e., that they are believing or judging that *p*. Objectifying seems to give the subject metarepresentation, but not ground-level conscious thinking. (Martínez-Manrique and Vicente, 2015, p. 4)

⁸ Thanks to Peter Langland-Hassan for suggesting this way of putting it.

⁹ Some theorists suggest that the two types of processing are supported by different mental systems or suites of systems, ‘System 1’ and ‘System 2’ (e.g., Evans & Over, 1996; Sloman, 1996; Stanovich, 1999, 2004). It is sometimes claimed that System 1 is evolutionarily ancient and largely shared with non-human animals, whereas System 2 is late-evolving and distinctively human (e.g., Stanovich, 2004).

low confidence in the intuitive response), when it provides a more reflective response which may override the intuitive one (e.g., Evans, 2006; Kahneman, 2011).

In outline at least, a dual-process view is well supported by experimental data from reasoning tasks. The hypothesis that behavioural control shifts between two qualitatively different processes neatly explains the effects of various experimental manipulations and the correlations observed between response patterns and individual differences in cognitive ability (for summaries, see the surveys cited earlier). For example, in syllogistic reasoning tasks, people have a tendency to endorse believable conclusions ('belief bias') instead of following instructions to assess logical validity, and this tendency can be increased or reduced by experimental manipulations (time pressure increases it, whereas strong explicit instructions reduce it; Evans & Curtis-Holmes, 2005; Evans et al., 1994).

Nevertheless, dual-process theories have attracted criticism (Keren & Schul, 2009; Kruglanski & Gigerenzer, 2011; Osman, 2004). They are often interpreted (and have sometimes been presented) as making implausibly strong claims, among them that the features ascribed to each process are all essential ones, that each type of processing is supported by a single neural system, and that cognitive biases are always the result of Type 1 processes. In recent years, dual-process theorists have responded by repudiating such strong claims and seeking to define a pared-down form of dual-process theory, which identifies the core difference between the two types of processing and allows for much more variation within each category (e.g., Evans, 2012; Evans and Stanovich, 2013). Evans identifies the core distinguishing feature as use of working memory — understood as a central, limited-capacity, attention-controlled memory store (Baddeley, 2007). Type 1 processes bypass working memory, whereas Type 2 processes load on it. As Evans notes, this explains the clusters of properties typically associated with each type of process. Processes that load on working memory will be serial, and they will typically be slower and of lower capacity than ones that do not, and display higher individual variability. (There is strong evidence that working memory capacity is correlated with measures of fluid intelligence; e.g., Colom et al., 2004; Kane et al., 2005.) Moreover, since attention is required to hold items in working memory, reasoning processes that use it will be relatively effortful and controlled in comparison to ones that do not, and the states involved will be globally broadcast and thus conscious (see Carruthers, 2015, pp. 82–8).

6. Type 2 thinking as activity

I am sympathetic to this pared-down dual-process approach, but I don't think it gets right to the heart of the matter. The fundamental difference between Type 1 and Type 2 reasoning, I suggest, is that the latter involves intentional action. I have made the case for this in earlier work, so I shall be brief here (see Frankish, 1998, 2004, 2009).¹⁰

¹⁰ For a similar construal of dual-process theory, see Carruthers 2006, 2009, 2015 — the last of which presents an impressively detailed case for an action-based view of reflective thought.

Most of our behaviour is spontaneous; we act without prior conscious thought. Think of driving, playing a sport, or holding a relaxed conversation. The behaviour involved is intelligent, guided by our knowledge, directed to our goals, and continually sensitive to incoming perceptual information, and complex information processing must be involved in selecting the right responses — when to brake, which shot to play, how to answer a question. However, this processing does not involve any activity on our part. We do not *do anything* in order to work out how to respond; all we do is press the brake, play the shot, answer the question. The reasoning processes involved in selecting the response are, I shall say, *autonomous* ones, in the sense that they are not under intentional control. It is true that we sometimes talk as if our spontaneous actions were the product of prior activity on our part. We say, for example, that we braked because we *realized* that the driver in front was slowing down. But this seems to be post hoc explanation or rationalization. (It may be objected that we should not assume that a process is not intentional just because it is not conscious. The range of intentional actions may be wider than we suppose, and may include mental actions, such as the control of attention (Carruthers, 2015). I agree, but there is a strong presumption that the processes that initiate spontaneous action are not themselves actions. They do not involve bodily movement or its motor precursors, and they could not *all* be intentional, under pain of regress.)

Yet there is also such a thing as deliberate, intentionally controlled reasoning. Think, for example, of doing complex arithmetic. If I am presented with a problem in long division, the answer does not come to me spontaneously. Rather, I have to *work it out*, by performing an explicit calculation. I take a pen, write down the figures in a certain format, and do various simpler calculations (whose answers *do* come to me spontaneously), adding further figures at each stage, until I complete the procedure and finally read off a sequence of figures as the answer. The actions collectively implement a procedure for solving the problem. I'll call this *intentional* reasoning, and I want to highlight some of its features.

First, the actions involved (writing the various figures) are intentional ones, which have belief-desire explanations. I decide to perform each step because I want to solve the overall problem and believe that the step is part of a procedure for solving it.

Second, the procedure involves episodes of autonomous reasoning. At each step, I must interpret the symbols I have written as posing various subproblems (given the truth of the ISA theory, the intended meaning of the symbols is not transparent to me), solve those subproblems, and decide which symbols to write next and where. And these processes will typically be autonomous ones, issuing in spontaneous behaviour. For example, if I need to subtract 9 from 17, I will just 'see' that the answer is 8. (Or, If I do need to do some intentional reasoning in order to solve a subproblem, in the form of marginal calculations, then the processes that guide and support *this* reasoning will be autonomous.) Thus, intentional reasoning is not *wholly* intentional, but guided and mediated by autonomous reasoning. Indeed, the purpose of the actions involved is, I suggest, precisely to break down the initial problem into simpler subproblems that can

be solved by autonomous processes. Intentional reasoning is a sort of deliberative mastication (chewing it over, as we say).

Third, my example of intentional reasoning involved the creation and manipulation of external symbols, but this is not essential to intentional reasoning. We can work things out ‘in our heads’, using visual or auditory images of the relevant symbols, held in working memory. Our working memory capacity is not sufficient to enable us to do long division in our heads, but we can tally a list of figures in this way, articulating the running total in inner speech. We may also be able to use other mental imagery in intentional reasoning, deliberately conjuring up images of objects and scenes as a way of helping us to evaluate different options (Carruthers, 2015, p. 159). (I assume here that we can actively form and manipulate mental imagery. We have already seen how some imagery may be formed and sustained by the mental rehearsal of action, and it is plausible to think that other forms of imagery can be actively generated, too, by intentionally directing attention; Carruthers, 2015, ch. 6.)

Fourth, the long division example involves executing a formal procedure for solving the problem, but not all intentional reasoning is like this. As I shall illustrate in the next section, we can employ informal, content-based procedures in intentional reasoning.

Now, my proposal is that it is the distinction between autonomous and intentional reasoning that lies at the core of dual-process theories. Type 1 processes are autonomous, whereas Type 2 processes are intentional. (Again, ‘intentional’ here means ‘involving intentional action’, not ‘wholly intentional’. Intentional processes have autonomous processes as parts.) In effect, this proposal takes the property of being controlled as the defining feature of Type 2 reasoning, specifying the relevant form of control as intentional.

This proposal coincides closely with Evans’s definition in terms of the use of working memory. For, at least when internalized, intentional reasoning involves forming, sustaining, and manipulating mental imagery, and will thus load heavily on working memory.¹¹ The proposal therefore explains all the features of Type 2 processing that the working memory proposal does — its seriality, relative slowness, lower capacity, effortfulness, individual variability, and the fact that it is, at least in part, conscious. Autonomous processes, on the other hand, do not require working memory and display a contrasting profile.¹²

Moreover, the proposal explains another feature of Type 2 processing that is not explained by the working memory proposal alone — namely, that Type 2 reasoning

¹¹ When external symbols are employed, however, intentional reasoning could in principle be done without use of working memory. Think of a bookkeeper doing routine calculations in a ledger. For short spells at least, they might work without attending to what they are doing, just as we can drive or read aloud without attention. Such unattended intentional reasoning might make little or no use of working memory and possess some of the features typically assigned to Type 1 processes, such as being fast and nonconscious.

¹² Might there not be a type of reasoning that loads on working memory but is not intentional — giving a triple-process theory? Nothing I have said rules out that possibility, but it is for the triple-process theorist to make that case that the data cannot be explained on the proposed dual-process approach. For a powerful case for the view that all reflective thinking is active, see Carruthers, 2015.

processes are *malleable*, capable of being shaped by imitation, tuition, and culture. As Carruthers has emphasized, this is exactly what we should expect if Type 2 reasoning is an activity, since our behaviour is malleable in just this way. We can learn new skills and regulate our activities in line with normative beliefs about how they ought to be conducted (Carruthers, 2009). This further explains why Type 2 processes are often the source of normative responding on reasoning tasks. Such responding typically involves following explicit task instructions or adhering to known normative standards, which we can constrain ourselves to do if we reason intentionally. (Of course, it does not follow that intentional reasoning will always yield the normatively correct result; people can acquire bad habits of reasoning and misunderstand or misapply instructions.)

On this view, then, dual-process theories assume a different aspect. The two kinds of reasoning are not supported by distinct neural systems, since all the resources that are available for autonomous, Type 1 reasoning can also be involved in supporting intentional, Type 2 reasoning. The latter involves harnessing Type 1 processes, together with working memory and motor systems, in the service of extended conscious problem-solving activities. We might think of the distinction as one of levels of organization: Type 2 reasoning forms a sort of ‘virtual’ system, realized in a range of neural sub-systems, and programmed with culturally transmitted skills and knowledge.¹³

The upshot of this is that there are strong grounds for regarding conscious reasoning as an activity, in which we produce and manipulate perceptually available symbols, or images of them. From this perspective, the format and activity views are not only compatible but complementary: natural language supplies a rich and flexible representational medium which can be used in active, conscious thinking, first overtly in private outer speech and then covertly in inner speech. (I shall suggest in the next section that this kind of private speech is itself an adaptation of a certain sort of social speech.) The use of inner speech for conscious thought is thus continuous with a use of outer speech, in line with the continuity thesis.

7. Speaking as thinking

Let us look at how speech can be used in intentional reasoning. I described such reasoning as deliberative mastication: the actions involved serve to decompose a complex problem into subproblems that can be solved by autonomous processes. The long division example illustrates this. Doing long division involves creating, manipulating, and responding to an external symbolic environment. We begin by writing out the figures in a certain format. We perceive the symbols and their arrangement, and autonomous processes interpret them as posing a simpler division problem, drawing both on general arithmetical knowledge and knowledge of this particular procedure. (This knowledge needn’t be declarative; some of it might be

¹³ In previous work, I have used the sub-personal/personal distinction to characterize these levels: Type 1 reasoning is reasoning that is done by neural subsystems, and Type 2 reasoning is reasoning that is done by people (Frankish, 2004, 2009). For the idea of a virtual system, see Dennett, 1991.

procedural, embedded in automatized skills in the manipulation of symbols.) Autonomous processes then generate a belief about the solution to this subproblem and a decision to write down further symbols expressing it. We write these symbols in the appropriate place, thereby modifying our symbolic environment. The process repeats until we have created a symbolic structure that we interpret as expressing a solution to the original problem (or we give up the search for one). So the process is a cyclical one: produce symbols, perceive them, interpret them as posing a subproblem (or a solution), form beliefs about the response to this subproblem, decide to produce further symbols expressing this response, and so on. The actions involved (writing the numerals) continually restructure the symbolic environment in ways that — if the procedure is sound — move one towards an overall solution.¹⁴

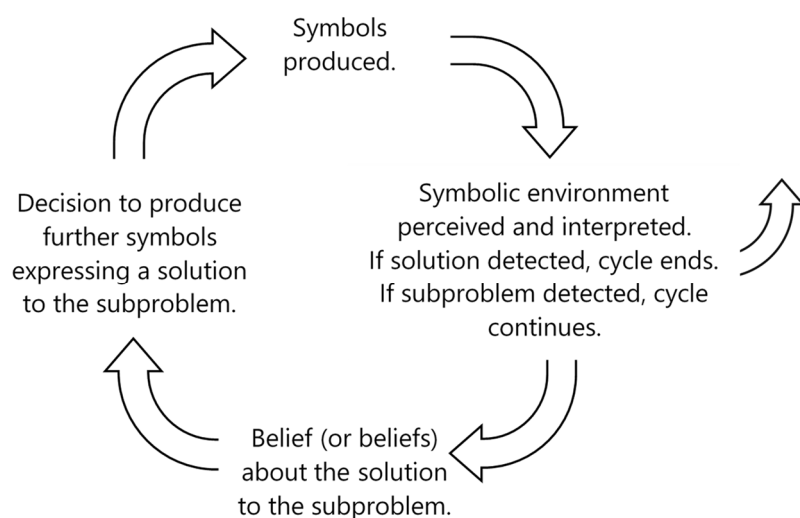


Figure 1: Intentional reasoning as a cyclical process

This model applies to speech-involving reasoning too. Most of us know some formal procedures for constructing arguments, either because we have been taught the relevant principles explicitly or because we have learned to imitate certain formal argumentative moves. Again, the process of applying these procedures has a cyclical structure: we start with a sentence, interpret it as a step in an argument, form a belief about the next step, add that sentence, and so on. Most speech-based reasoning, however, does not employ formal principles, relying instead on (autonomous, nonconscious) judgments of relevance or evidential support, or on simple associations. But it still has the same basic cyclical structure. Here is an example. Suppose I have been invited to a party with colleagues from work. I don't find myself strongly disposed to respond one way or the other, but I need to give an answer, so I engage in intentional reasoning. (If the solution won't come to me, we might say, I must go to it.) Again, I need some way of breaking

¹⁴ For extended development of the idea that conscious ('System 2') reasoning consists of cycles of mental rehearsal and nonconscious, modular processing, see Carruthers, 2006, chs.4-7.

down the problem. I begin by questioning myself to try to elicit an evaluative reaction or piece of relevant information, asking, 'Do I really want to go?', 'What will it be like?', or something similar. I hear my own utterance, my language comprehension system interprets it, and its content is globally broadcast to other mental subsystems. My mindreading faculty interprets me as requesting information about the party or an evaluation of it, and further automatic processes throw up the prediction, based on experience, that Henry will be there. This message is selected for expression (perhaps winning a competition among other candidate messages; see Carruthers, this volume; Dennett, 1991), and I utter the words: 'Henry will probably be there'. Again, this utterance is heard and interpreted. Though it does not have the form of a question, in the context it is interpreted as posing further subproblems: Do I want to meet Henry? What will happen when I meet him? Again, a response is selected and articulated: 'He'll want to talk about the budget cuts'. This in turn is heard and interpreted as posing the problem of whether I want to talk about the budget cuts. My affective response — let us suppose — is strongly negative, and I conclude by uttering, 'I can't face that; I won't go'.

We might use a similar strategy in theoretical reasoning. If invited to give an opinion of a proposed economic policy, I might ask myself, 'Would that work?', prompting autonomous processes to produce an utterance articulating a likely consequence of the policy, which might then prompt a further utterance expressing a reaction to that consequence, and so on, continuing the process until I find myself inclined to express a judgement on the policy as a whole.

These examples are formally similar to the long division one. My autonomous processes don't deliver a decision about the party or an evaluation of the economic policy, just as they don't produce the answer to a long division problem. So, I break the problems down into subproblems which my autonomous processes can solve. I immediately realize that Henry will want to talk about the budget cuts, just as I immediately see that 17 minus 9 is 8. By chaining these subproblems together in the right way, I reach a solution to the original problem.

I suggest that this is a plausible schema for speech-involving reasoning, conducted in outer, private speech. Of course, the examples are schematic, and real-life examples will often be less explicit and less straightforward. Some of the utterances might be abbreviated or even replaced by visual imagery (instead of starting off with a question, I might try to imagine the party). The sequence of utterances might take unproductive twists and turns, run into a dead end, or just peter out. Moreover, the strategy may yield wrong or suboptimal solutions. The responses and evaluations evoked by one's utterances may be heavily influenced by associations and contextual factors that are not relevant to the problem. Good reasoners will be cautious in adopting conclusions they come to and will persist with the reflective process until they feel a high level of confidence in the result. But the activity is a hugely valuable one all the same, enabling us to tackle complex and unfamiliar problems which do not evoke immediate, intuitive responses.

It is not hard to see how we might learn to engage in deliberative activity like this. For it involves questioning and prompting ourselves, and we do a similar thing in social

speech, questioning and prompting other people — children especially — to help them arrive at a decision or to form an opinion. If my young daughter can't decide whether or not to accept an invitation to Emily's party, I will ask her, 'Do you like Emily?', 'Who else will be there?', 'Do you like playing with Emily's friends?', and so on, providing linguistic stimuli that help her to get a purchase on the problem and come to a decision. Children are routinely exposed to such strategies, and I suggest that Type 2 reasoning emerges when they begin to apply them to themselves, first in private speech and later in inner speech. (This is, in a sense, an inversion of the process envisaged by defenders of the extended mind thesis, in which external artefacts come to implement cognitive states and processes previously located within the head (e.g. Clark and Chalmers, 1998). On the view proposed here, by contrast, the opposite happens. Conscious thinking is initially externalized, dependent on outer speech, and inner speech enables us to internalize it, forming a *contracted* mind.)

It may be objected here that the activities I have been describing do not deserve the title of reasoning or thinking. The utterances are not playing a direct cognitive role but merely a communicative one, passing along responses to other systems for further processing. There are several points to make in response to this. First, even if the utterances were merely passing information, they would still be essential components in a larger, temporally extended reasoning process, since there may be no internal channels to pass the information. There is a strong case for thinking that the brain has a modular architecture, with only limited internal access channels between modules (e.g., Carruthers, 2006). If that is right, then a response may need to be globally broadcast in order to reach the subsystems required for its further processing. And if it is only sensory information that is globally broadcast, as the ISA theory holds, then it will need to be represented in some sensory format in order for this to happen.

Second, it is not the case that the utterances are simply passing on information. They are passing it on *in a certain context* — in the context of a symbolic environment recording an extended engagement with a larger problem. Each utterance is interpreted in the context of preceding utterances. Take the information that Henry will be at the party. All sorts of things follow from this — that Henry won't be at home, that there will be at least one person at the party, that the party goers will not all be female, and so on. But when heard in the context of a series of utterances about whether *I want to go to the party*, it takes on a special significance: How will Henry's being there affect my enjoyment of the party? What will happen when I meet Henry?

Third, producing an utterance has a *selectional* effect. At each stage of the cycle there are many possible responses to the particular subproblem posed. Autonomous processes may generate a number of these and compete to get them articulated. Since only one response can win, be articulated, and go through to the next round, the process of articulation profoundly shapes the overall reasoning process. Different choices of utterance may take the process in completely different directions. By its nature as a selective, serial process, intentional reasoning carves out a specific route through the deliberative territory, which we would not otherwise have taken.

8. Speaking as judging and deciding

We have seen how speech can enable us to construct conscious trains of reasoning leading to a conclusion. However, conscious thinking involves more than this: it also involves making judgements and decisions. When my train of reasoning leads me to a conclusion — that I shouldn't go to the party or that the economic policy will work — I need to form the corresponding mental attitude.¹⁵ I need to decide not to go to the party or form the belief that the policy will work. But how do I do this? Just *saying*, 'I won't go to the party', or 'The policy will work' isn't enough.

One possibility is that we cannot make conscious judgements and decisions. All we can do is utter a conclusion and hope that nonconscious processes form the corresponding decision or judgement — that the suggestion 'takes'. It is not implausible to think that uttered conclusions might have this effect. We often respond in this way to assertions and advice from other people. If someone tells me that something is true or advises me to do something, then, if I think they are trustworthy and well-informed, I (or, rather, my autonomous mental systems) may form the corresponding belief or decision. I am even more likely to trust assertions and advice from myself, especially if I believe they are the product of a train of reasoning aimed at finding the best conclusion. However, if this were the only way in which our conscious reasoning could influence our attitudes, then conscious thought would be rather fragile. I couldn't be confident that I would respond to my own assertions and advice, any more than I could be confident that other people would.

In fact, I think there is another, more reliable way in which conscious reasoning can affect our attitudes. Indeed, I have argued elsewhere that in uttering conclusions we can form distinctively conscious attitudes — beliefs, desires, decisions, and intentions — which may differ from our nonconscious ones (Frankish, 2004). The proposal appeals to another use of outer speech. Think about the kinds of speech act I might perform by saying 'I won't go to the party'. I might be making a prediction, or entertaining a hypothesis, or expressing a fear. But there's something else I might be doing: I might be making a *promise* or *commitment*. Suppose my partner doesn't want me to go to the party because we have to leave early the next morning on a family trip. If I tell her that I won't go, then I have committed myself to not going, and if she later discovers that I have gone, she will rightly be annoyed with me for not keeping my word. Expressions of desire or intention can also serve as public commitments. Think about asking a child what they want from an ice cream stall. We are not asking them to introspect — to examine their minds to see if they find a ready-formed preference or decision. (If the ISA theory is right, they will not be able to do that anyway.) Rather, we are asking them to *make up their mind* — to commit themselves to an option. And once they have done so, we shall expect them to stick to it and not capriciously reject the item for which they have plumped. (And if they do reject it, we shall blame them for their inconstancy, not

¹⁵ In this chapter, I use 'attitude' in the philosophical sense, to mean a *propositional* attitude — believing, desiring, deciding, wondering etc. This contrasts with social psychologists' use of the term for an overall evaluation of something.

excuse them as having made an error of introspection. To achieve constancy, they don't need to introspect but simply to remember their avowed choice and stick to it.) Something similar can happen when we are asked our opinion about something. Often, it is not simply a matter of reporting what we believe. We may have no settled opinion on the matter. But under pressure to state an opinion we may opt for a view and commit to it. And once we have done this, we shall be expected to stick to the opinion — at least for a while.

Some social utterances, then, express a commitment to an action, opinion, or preference. And I suggest that some private utterances have a similar function. Having been schooled in making commitments of this kind to others, we start to make them to ourselves — committing ourselves to regulating our future activities, including our intentional reasoning, in line with the choice or view expressed. Thus, in saying that I won't go to the party I commit myself to not going to the party and to taking my not going as a fixed point in my future conscious reasoning — making further commitments required by my not going and refraining from making commitments incompatible with my not going. Similarly, in telling myself that the economic policy will work, I commit myself to regulating my behaviour in line with that view — asserting it (when appropriate), defending it, and treating it as true in my intentional reasoning (taking it as a premise, rejecting conclusions incompatible with it, and so on). And such commitments, I suggest, constitute a distinct kind of mental attitude — conscious decisions, conscious beliefs, conscious desires — which may differ from the nonconscious ones formed by autonomous processes. I may have consciously decided not to go to the party (committed myself to that course) or formed the conscious belief that the economic policy will work (committed myself to treating the claim as true) even though my nonconscious, autonomous processes have not formed that decision or that belief.

It may be objected that on this view conscious thought is still dependent on nonconscious processes for its efficacy. In general, an utterance constitutes a commitment only if the participants understand it as one. Similarly, a self-directed utterance will constitute a commitment only if it is heard and interpreted as one by nonconscious, autonomous processes. (Thus, even if I can form a conscious attitude without forming the corresponding nonconscious attitude, I cannot form it without forming *any* nonconscious attitude; I must form the nonconscious belief that I have made an appropriate commitment.) Moreover, if the commitment is to be effective in guiding my behaviour, I must have a general desire to act on my commitments. So, as in the previous scenario we considered, in which conscious thoughts served only as self-assertions and self-advice, I cannot be sure that my conscious thoughts will have the appropriate effects.

This is true, but there are reasons for thinking that psychological commitments of the kind described will typically be interpreted correctly and acted upon. First, there will be strong contextual cues for the interpretation. A commitment-expressing utterance comes at the conclusion of an episode of intentional reasoning and expresses a solution to the problem one has been considering, and it may be expressed in a tone

of resolution and accompanied by a feeling of satisfaction. Second, it is likely that we have a general desire to honour our self-directed commitments. Social commitments enable us to coordinate our activities with those of others. There are huge advantages to this, and consequently strong social pressure on us to honour our commitments (through fear of censure, ostracization, and so on). And this pressure is likely to transfer to our self-directed commitments. For such commitments often acquire a social dimension. Committing oneself to a view, preference, or plan involves being prepared to declare it, defend it, and act upon it, and hence to incur a social commitment to it too. People expect us to stick to our avowed attitudes and plans, especially if they have modified their own behaviour to take account of them. Moreover, there are psychological advantages to honouring our private, self-directed commitments. Such commitments serve a similar function to social ones — they help us to coordinate our own activities over time. Having made a conscious commitment to an opinion, preference, or course of action, I shall expect myself to stick to it — to constrain myself to reasoning and acting accordingly, even if my corresponding first-order beliefs and desires fluctuate in strength. (If I have committed myself to not going to the party, then I shall expect myself not to go, even if I am subsequently tempted to do so.) Commitments thus play an important stabilizing role in our lives, giving us a new level of self-awareness and self-control, and enabling us to resist whims and passing temptations. We have, then, strong reason to honour our self-directed commitments — to be *strong-willed*. (For more on strength of will and self-directed commitment, see Frankish, 2016.)

Moreover, there is an important difference between the self-assertion/advice view and the self-commitment view. On the former, autonomous processes need to do two things: (a) correctly interpret the utterance (as assertion or advice), and (b) form the appropriate belief or decision. But on the commitment view, only one step is needed. If I interpret myself as having made a certain commitment, then, effectively, I *have* made it. Believing yourself to be committed to a certain course of action is sufficient for being committed to it. Commitment-beliefs are self-fulfilling.

There are two more objections to consider, parallel to the ones I raised against the self-communication view of conscious thinking discussed in section 4. The first objection is that commitment-expressing utterances cannot count as conscious thoughts, since we do not have direct, non-inferential knowledge of them. Like self-communicative utterances, they have to be interpreted in order for us to recover their meaning. I respond by appealing to the point just made, that beliefs about our commitments are self-fulfilling. In the case of self-communications, there can be a gap between the attitudes I interpret myself as expressing and the ones I really have. I can misinterpret myself as expressing a belief that I do not in fact have. But in the case of self-commitment, there can be no such mismatch. If I interpret myself as expressing a commitment to a certain action, then I thereby become committed to the action. The belief constitutes the commitment. Thus, even if an utterance wasn't originally intended to express a commitment, it may retrospectively acquire that status in virtue of its subsequent interpretation. The moral of this is that we have a direct, *constitutive*

authority about our commitment-expressing utterances, which justifies the identification of conscious decisions and judgements with such utterances (Frankish, 2004, ch. 8).

The second objection, which has been forcefully pressed by Carruthers, concerns the cognitive role of self-commitments (see Carruthers, 2011, ch. 4, 2013, 2015, ch. 7). In essence, the objection is this. A decision to do something *settles* the matter. If I decide to buy a new car, then I may subsequently reason about how to implement this decision (which kind of car to buy, where to buy from and so on), but (unless I change my mind) I shall not engage in further reasoning about whether or not to buy a new car. However, a commitment to doing something does not terminate reasoning in this way. The belief that I am committed to buying a new car does not settle that I shall buy a new car. Further reasoning is required, in which the belief interacts with the desire to honour my commitments to produce a decision to buy a new car. (If that desire is chronically weak, I may never get to the car showroom.) Similarly, a judgement settles how I treat the judgement's content. If I judge that inflation will rise soon, then I become immediately disposed to assert that inflation will rise soon, defend the view that inflation will rise soon, take it as a premise that inflation will rise soon, and so on. But a commitment to treating it as true that inflation will rise soon does not immediately dispose me to do those things. Again, mediating reasoning is needed, involving a desire to honour my commitments.

This argument deserves far more discussion than I can give it here, but I shall indicate the line of reply I favour (for a more detailed reply, see Frankish, 2012). It involves taking seriously the idea mentioned earlier, that conscious, intentional reasoning and its associated attitudes form a distinct level of mentality, which is realized in lower-level nonconscious processes and attitudes (in cycles of autonomous reasoning and nonconscious beliefs about commitments). From this perspective, claims about an attitude's functional role should, plausibly, be relativized to a level. A conscious decision should terminate reasoning *at the conscious level*, and a conscious judgement should fix how we treat its content *in conscious reasoning*. And on the view proposed, this will be so. Having formed the conscious decision to buy a new car (that is, committed myself to that course), I shall cease to engage in conscious, intentional reasoning about whether to buy a car, and may head off to the showroom without further conscious thought. Similarly, having judged that inflation will rise soon, I shall treat that proposition as true in subsequent conscious reasoning, and assert it and defend it without further conscious reasoning about its truth. Of course, my disposition to do these things depends on further reasoning occurring at the nonconscious level — reasoning about my commitments and my desire to honour them — but these lower-level processes do not form part of the functional role of the conscious attitudes they realize.

It may be asked why we should adopt this perspective. Why not take a single-level view, on which all our attitudes are nonconscious but we have higher-order beliefs about our commitments to opinions, preferences, and actions, as well first-order beliefs, desires, and decisions. Perhaps there is a theoretical case for taking such a view (see the previously cited works by Carruthers). But in practice at least, it is natural to think of

the commitments as constituting first-order attitudes of a distinct kind. (We might call them *virtual* attitudes.) For, being conscious, these attitudes are specially salient to us, and by treating them as first-order ones we can easily predict our own behaviour and coordinate it over time. Compare the social commitments on which self-directed commitments of this kind are modelled. When I urge my daughter to make up her mind, I am urging her to commit to a view, a preference, or a course of action, but I shall think of her response simply as expressing a belief, desire, or decision, on which I shall expect her to act directly. If she regulates her behaviour in line with her commitments, she will behave in much the same way as if she were acting directly on corresponding first-order attitudes, and it will be far simpler to think of her as doing that. Forming conscious attitudes, I suggest, involves adopting a similar attitude towards oneself, prompting oneself to express a commitment and then assuming that one will behave as if one had the corresponding first-order attitude.

Conclusion

There are good reasons for thinking of inner speech both as a medium for conscious thought and as an activity whose functions are continuous with those of outer speech. These apparently conflicting views can be reconciled by adopting a version of dual-process theory on which conscious thinking is an intentional activity, rooted in certain social uses of language. Conscious thinking involves the self-application of interpersonal practices of questioning and prompting, and conscious attitudes are a self-directed version of social commitments. The activities involved are initially conducted in outer speech, but inner speech allows us to perform them covertly, internalizing outer thought.¹⁶

References

- Alderson-Day, B. & Fernyhough, C. (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin* 141(5), 931–65.
- Baars, B. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.
- Baddeley, A. D. (2007). *Working memory, thought and action*. Oxford: Oxford University Press.
- Bentall, R. P. (1990). The illusion of reality: A review and integration of psychological research on hallucinations. *Psychological Bulletin*, 107(1), 82–95.

¹⁶ An earlier version of this paper was presented at a workshop, Inner Speech: Theories and Models, organized by the Universities of Granada and the Basque Country, and held in Granada, Spain, in July 2015. I am grateful to all the participants in the workshop for their questions, comments, and suggestions. Thanks are also due to Peter Carruthers, Eileen Frankish, Maria Kasmirli, and the editors of this volume for comments, advice, and assistance.

- Carruthers, P. (2006). *The architecture of the mind: Massive modularity and the flexibility of thought*. Oxford: Oxford University Press.
- Carruthers, P. (2009). An architecture for dual reasoning. In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 109–27). Oxford: Oxford University Press.
- Carruthers, P. (2011). *The opacity of mind: An integrative theory of self-knowledge*. Oxford: Oxford University Press.
- Carruthers, P. (2013). On knowing your own beliefs: A representationalist account. In N. Nottelmann (Ed.), *New essays on belief* (pp. 145–65). Basingstoke: Palgrave Macmillan.
- Carruthers, P. (2015). *The centred mind: What the science of working memory shows us about the nature of human thought*. Oxford: Oxford University Press.
- Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.
- Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P. (2004). Working memory is (almost) perfectly predicted by *g*. *Intelligence*, 32, 277–96.
- Dennett, D. C. (1991). *Consciousness explained*. London: Allen Lane.
- Driskell, J. E., Copper, C., & Moran, A. (1994). Does mental practice enhance performance? *Journal of Applied Psychology*, 79(4): 481–92.
- Evans, J. St. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, 13(3), 378–95.
- Evans, J. St. B. T. (2011). *Thinking Twice: Two minds in one brain*. Oxford: Oxford University Press.
- Evans, J. St. B. T. (2012). Dual-process theories of deductive reasoning: Facts and fallacies. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 115–33). Oxford: Oxford University Press.
- Evans, J. St. B. T. & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, 11(4), 382–9.
- Evans, J. St. B. T., Newstead, S. E., Allen, J. L., & Pollard, P. (1994). Debiasing by instruction: The case of belief bias. *European Journal of Cognitive Psychology*, 6(3), 263–85.
- Evans, J. St. B. T. & Over, D. E. (1996). *Rationality and reasoning*. Hove: Psychology Press.
- Evans, J. St. B. T. & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3), 223–41.
- Frankish, K. (1998). Natural language and virtual belief. In P. Carruthers & J. Boucher (Eds.), *Language and thought: interdisciplinary themes* (pp. 248–69). Cambridge: Cambridge University Press.
- Frankish, K. (2004). *Mind and supermind*. Cambridge: Cambridge University Press.
- Frankish, K. (2009). Systems and levels: Dual-system theories and the personal–subpersonal distinction. In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 89–107). Oxford: Oxford University Press.

- Frankish, K. (2010). Dual-process and dual-system theories of reasoning. *Philosophy Compass*, 5(10), 914–26.
- Frankish, K. (2012). Dual systems and dual attitudes. *Mind & Society*, 11(1), 41–51.
- Frankish, K. (2016) Playing double: Implicit bias, dual levels, and self-control. In M. Brownstein & J. Saul (Eds.), *Implicit bias and philosophy volume I: Metaphysics and epistemology* (pp.23–46). Oxford: Oxford University Press.
- Frankish, K. & Evans, J. St. B. T. (2009). The duality of mind: An historical perspective. In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 1–29). Oxford: Oxford University Press.
- Frith, C. (1992). *The cognitive neuropsychology of schizophrenia*. Hove: Psychology Press.
- Jackendoff, R. (1996). How language helps us think. *Pragmatics & Cognition*, 4(1), 1–34.
- Kahneman, D. (2011). *Thinking, fast and slow*. London: Allen Lane.
- Kane, M., Hambrick, D., & Conway, A. (2005). Working memory capacity and fluid intelligence are strongly related constructs. *Psychological Bulletin*, 131(1), 66–71.
- Keren, G. & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, 4(6), 533–50.
- Kruglanski, A. W. & Gigerenzer, G. (2011). Intuitive and deliberative judgements are based on common principles. *Psychological Review*, 118(1), 97–109.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Martínez-Manrique, F. & Vicente, A. (2015). The activity view of inner speech. *Frontiers in Psychology*, 6, Article 232, 1–13.
- Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, 11(6), 988–1010.
- Prinz, J. (2011). The sensory basis of cognitive phenomenology. In T. Bayne & M. Montague (Eds.), *Cognitive phenomenology* (pp. 174–96). Oxford: Oxford University Press.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119(1), 3–22.
- Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Stanovich, K. E. (2004). *The robot's rebellion*. Chicago: The University of Chicago Press.
- Winsler, A., Fernyhough, C., & Montero, I. (Eds.) (2009). *Private speech, executive functioning, and the development of verbal self-regulation*. Cambridge: Cambridge University Press.
- Winsler, A. & Naglieri, J. (2003). Overt and covert verbal problem-solving strategies: Developmental trends in use, awareness, and relations with task performance in children aged 5 to 17. *Child Development*, 74(3), 659–78.