THOUGHT EXPERIMENTS AND MENTAL SIMULATIONS

ABSTRACT

Thought experiments have a mysterious way of informing us about the world, apparently without examining it, yet with a great degree of certainty. It is tempting to try to explain this capacity by making use of the idea that in thought experiments, the mind somehow simulates the processes about which it reaches conclusions. Here, I test this idea. I argue that when they predict the outcomes of hypothetical *physical* situations, thought experiments cannot simulate physical processes. They use mental models, which should not be confused with process-driven simulations. A convincing case can be made that thought experiments about hypothetical *mental* processes are mental simulations. Concerning *moral* thought experiments, I argue that construing them as simulations of mental processes favours certain moral theories over others. The scope of mental simulation in thought experiments is primarily limited by the constraint of relevant similarity on source and target processes: on one hand, this constraint disqualifies thought from simulating external natural processes; on the other hand, it is a source of epistemic bias in moral thought experiments. In view of these results, I conclude that thought experiments and mental simulations cannot be assimilated as means of acquiring knowledge.

SIMULATIONS, THOUGHTS, AND NATURAL PROCESSES

There are two ways in which thought experiments could be mental simulations, depending on what kind of process is being simulated. Some thought experiments could be mental processes simulating *non-mental* processes; others could be mental processes simulating *mental* target-processes. The second category would contain thought experiments which seek to acquaint us imaginatively with moral dilemmas, whether by

engaging our reasoning or by replicating the epistemic conditions under which moral sentiments may be aroused. Such thought experiments could be called 'mental-mental' simulations (mental source-processes, mental target-processes). Physical thought experiments, on the other hand, would be in the first category: they are conducted by means of processes (thoughts) which do not occur in their targets (external physical events), and could be termed 'mental-physical' simulations.

A standard example of simulation is testing a model airplane in a wind-tunnel (a *physical*-physical simulation). Although the variables the simulation is about do not appear in the simulation, as is also the case in thought experiments, we nevertheless, unlike in thought experiments, delegate to natural processes and not to thoughts the task of determining the outcome of the natural target process. Accordingly, when we have to simulate mental target processes, simulations are conducted by mental processes. This is consistent with a central thesis of the simulation theory of mind: there has to be some 'relevant similarity' (Davies 1998, Stone 1996, Stich and Nichols 1992) between source and target processes for a simulation to be possible in the first place. Since mental-mental simulations experiment *with* thoughts *about* thoughts, they are not substitutes for experimentation but genuine experiments – a status not normally granted thought experiments. But now, mental-*physical* simulations, which is what physical thought experiments would be if they were simulations, would appear to be an oddity for simulation theory: thought experiments are about physical processes but they are thought-conducted, so how can they be *simulations*?

Nevertheless, there have been several attempts (Gendler 2004, Gooding 1993, Nersessian 2002, 1999) to explain both the mental execution of thought experiments, and the principles underpinning their epistemological validity, by using some concept or other of simulation. These authors exploit a hunch already present in Ernst Mach, who claims that during physical thought experimentation our thoughts 'mimic' natural processes (see section 3). The concept of simulation, apart from requiring similarities between source and target processes, also implies *dissimilarities* – the dissimilarities between the variables appearing in the simulation and those whose behaviour the simulation is

intended to predict. It is by exploiting this aspect of simulation that theories of thought experimentation attempt to explain how thought can inform us of the workings of something as dissimilar from it as physical processes. At the same time, simulation theory itself offers some encouragement to such attempts, for it too sometimes seeks to bridge the gap between dissimilar processes or phenomena by defining special forms of epistemological access. For example, some simulationists describe mental visualization as a substitute for seeing (Currie 1995, Nersessian 2002; see also Walton 1990, Ch. 8), which is potentially useful for bridging the thought/observation gap in physical thought experiments. Others describe imagining – construed not as a form of mental representation, but as a special mental attitude adopted *towards* mental representations – as a substitute for belief (Nichols 2004, Walton 1990, Currie and Ravenscroft 2003). This is useful for analyzing thought experiments which seek to reconstruct the cognitive underpinnings of motivational and emotional states.

Apart from such explicit appeals to simulation, any theory which claims that thought experiments are not reducible to bare arguments also potentially leaves room for simulation to play some role, since something else has to stand in for arguments according to such theories. Besides, in certain respects, thought experiments and mental simulations are quite similar. Both are mental activities in which (a) we mentally represent a hypothetical or a counterfactual set of circumstances, and (b) we mentally process that *representation* in order to reach a prediction about how the set of circumstances it represents would behave. In each case it is required that without conditions actually obtaining as they would in an experiment, we can nevertheless, by using only our mental resources, make a prediction about those conditions.

Of course, the mental process by which we reach a prediction about the outcome of a system's physical state may be *neither* an argument *nor* a simulation. It may be an intuition of some as yet unspecified kind, or some (yet to be specified) use of prior implicit knowledge. My purpose here is precisely to clarify these options and choose between them by distinguishing the respective roles of: (a) process-driven simulation, (b) mental modelling, (c) visualization, (d) implicit knowledge, (e) induction, and (f)

inference. Unless this is done, the term ‘simulation’ (like ‘intuition’) will be used in as broad a sense as possible (covering (a) to (e) in the preceding list), and will merely serve to re-baptize, without dispelling, the mystery of how thought experiments can gain access to the workings of the thought-transcendent world.

THE RELEVANT SIMILARITY CONSTRAINT

The relevant similarity constraint on source and target processes applies only to *process-driven* simulations, or to simulations to the extent that they are process-driven (Davies 1998, Stone 1996, Stich and Nichols 1992). This leaves room for a concept of *theory-driven* mental simulation, which does not require relevant similarity because it merely involves application of prior theoretical knowledge to what we are imagining. This would occur if, for instance, in Stevinus’s thought experiment, a mental representation of the chain revolving perpetually did not conflict with any implicit knowledge (such as domain-specific know-how manifested in sensorimotor responses), and we had instead to reach the prediction that the chain does not revolve perpetually by applying an acquired theory that there is no perpetual motion. The view that thought experiments are *theory-driven* simulations would be compatible with Norton’s (2004) view that thought experiments are ‘merely picturesque arguments’ and that the knowledge of the natural world afforded by thought experiments comes ‘from premises introduced explicitly or tacitly into the thought experiment’ which is then ‘transformed, usually tacitly, through deductive or inductive argumentation to give the final result.’ Similar reservations about the possibility of process-driven simulation (and support for theory-drivenness) are expressed by Dennett even for *mental* targets during the activity of mindreading:

> If I make believe I am a suspension bridge and wonder what I will do when the wind blows, what ‘comes to me’ in my make-believe state depends on how sophisticated my knowledge is of the physics and engineering of suspension bridges. Why should my making believe I have your beliefs be any different? (Dennett 1989, 102)

If there was no relevant similarity between target and source processes in the putative simulations of thought-processes, thought-simulations would have to proceed by applying generalizations about situations to predict human behaviour. But against Dennett simulationists argue, plausibly enough, that there *is* relevant similarity, because some mental processes 'operate in just the same way when we imagine being in a particular situation as they would if we were really in that situation' (Davies and Stone 2000). Indeed, Dennett's own theory does not suppose that mindreading proceeds by treating individual cases of practical reasoning by applying acquired theoretical generalizations – something which would imply a strange notion of practical rationality. It instead holds that mindreading proceeds by applying normative rational constraints. This has given rise to Jane Heal's simulationist interpretation of Dennett's normative intentional stance (Heal 1998; this is the sense in which I apply the concept of simulation in sections 4 and 5 below). If the target mind uses counterfactual reasoning and applies some normative concept of belief to reach behaviour from beliefs and desires, my own mind will use the same processes to reach imagined behaviours from suppositions – because, and to the extent that, we can assume that my own counterfactual reasoning, concept of belief, semantic constraints, and so forth, are *relevantly similar* to the target subject's. I use my own mind to see what would happen in another mind.

This idealized description of mental simulation in mindreading shows what a clear case of relevant similarity between mental processes and targets *would* look like, and on what basis mental simulations *could* succeed. It shows that a mental activity is substantially a simulation only if it is process-driven, that is, *only if it simulates the processes by which the target goes from an initial state to a subsequent state*.

This condition is not met by everything we call a simulation in other senses of the term. A mental representation may be said to simulate what it represents in the weaker sense that it resembles or replicates spatial and temporal relations between the parts of the target, while at the same time, the *way* we reach predictions about the target's transition from one state to another is not by simulating the referent's processes. There is certainly no prima facie similarity between thought processes and processes which determine the

physical states of external objects. This is implicit in existing formulations of the epistemological problem of thought experiments: 'Thought experiments are supposed to give us knowledge of the natural world. From where does this knowledge come?' (Norton 2004), 'How is it possible to learn apparently new things about nature without new empirical data?' (Brown 2007). For, where mental simulation *does* work, in the mental simulation of thought-processes, two kinds of reply to these questions spring immediately to mind. (1) We learn new things about nature by using our thoughts because we experiment with our brains ('thoughts') to learn about other brains ('nature'). (2) We learn something about mental phenomena by experimenting with mental phenomena. In either case, the reply is available because it is in principle possible to claim some form of similarity between target and source processes. The problem is that relevant similarity does not seem to obtain other than in thought simulations of thought processes: how can experimentation processes which are not designed for physical-level predictions and explanations, but for mental predictions and explanations, make physical-level predictions and still be experiments?

PHYSICAL THOUGHT EXPERIMENTS

Attempts to respond to this problem in theories of thought experiments can be described as theses about *special epistemological access*. By this, I refer to a growing set of claims concerning the role in thought experimentation of various components of simulation theory, especially visualization and perceptual-motor activity in the manipulation of mental images, as well as to forms of non-propositional knowing, knowing-how, domain-specificity, or modularity.

Such appeals can already be found in Mach's concept of a form of unarticulated 'instinctive knowledge' derived from the observation of natural processes. Two traits of this instinctive knowledge of physical processes described by Mach are especially relevant to our purposes: (i) it 'exists in absolute independence of our participation'; and (ii) it is 'imprinted' 'in our percepts and ideas, which, then, in their turn, mimic the

processes of nature' (Ernst Mach, *The science of mechanics*, quoted by Sorensen 1992, 54, 51). The first characteristic implies that the implicit knowledge enables us to have certain expectations about the outcomes of physical processes as a result of processes which occur spontaneously, as autonomous events independent of our agency. This suggests that the ability which leads us to expect certain outcomes from imagined physical processes is sub-doxastic and non-inferential. If this is so, then the predictions drawn from thought experiments cannot result from the application of laws or other generalizations, and therefore is not theory-driven but process-driven. What remains to be seen, however, is whether this process is a simulation in the first place.

Mach's claim that perceptions *mimic* the processes of nature gives this impression. The claim can at best be cashed out as meaning that certain mental representations trigger affordance-based, domain-specific, perhaps modular, reactions to physical processes. But the 'mimicry' or replication in question will not be a replication of target physical process by cognitive or brain processes. When we perform Stevinus's thought experiment, not only do we not have a scaled-down prism-and-chain system in our heads, but we *cannot* have anything relevantly similar in respect of causal properties to the prism-chain system: for if the model were relevantly similar, then it would be nomologically impossible for the chain-analog to slide around the prism-analog in the way we can *represent* it as doing.

So Mach's concept of mimicry cannot imply a replication of the structural properties of the target system. What, then, is intended by the use of the concept? Surely, the fact that *just as* before a real chain on a prism we would be surprised to see it slide, we will be *similarly* be surprised before a spatial mental representation *of* a chain sliding around a prism. But all that *this* means is that our affordance-based dispositions towards physical objects can be triggered by imagining situations as well as by seeing situations. The replication or similarity relation obtains only between our *responses* to *F*s and our *responses* to mental representations of *F*s, not between mental processes and *F*s. These responses are enabled jointly by the mental representation, taken as an initiating cause, and by domain-specific know-how construed as a disposition. The initiating cause (the spatial mental representation) bears a perceptual similarity to the target, but neither the

inner cause (the implicit knowledge of the target process), nor the manifestation of the disposition (the expectation that Stevinus's chain will not revolve), bears any similarity to the target process. Since it is the inner cause which determines what imagined outcomes, similarity need not play any role in determining which imagined outcome we will accept. The similarity between visualizing and perceiving the target is due to preservation of spatial and topological features. But even iconic representations, as in fact Plato pointed out, bear only a *recognitional* similarity to their depicta, not a similarity in respect of observer-independent, world-to-world causal properties. Mach's example actually brings clearly into focus the fact that there cannot be a replication of target-properties in the mind or the brain, because the thought experiment affords us knowledge by mentally visualizing a physically impossible, but logically possible, event – something which can be done only by *avoiding* relevant similarity, that is, representationally.

An author who applies Mach's ideas by using current theories of psychology and philosophy of mind is Tamar Szabo Gendler. It is worth comparing the question that Gendler formulates in this connection with the question formulated here – namely, how experimentation processes designed for mental-level predictions can be expected to make physical-level predictions. Gendler asks 'how contemplation of an imaginary scenario can lead to new knowledge about contingent features of the natural world', and, significantly, sees this question as 'a special case of a more general one, namely how any *nonperceptual* capacity can lead to new knowledge about (nonstipulated) contingent features of reality' (Gendler 1152; italics added). This sets the course for her reply, which assimilates thought experimentation, construed as 'reasoning about an imaginary scenario', to the perceptual observation involved in ordinary experiments. The reply draws on the idea that affordance-based sense, proprioception, and manipulation of mental images generate new information in a non-argumentative way to acquire 'knowledge of contingent features of the natural world'. Gendler illustrates such reasoning with examples such as trying to figure out how many elephants can fit into a remembered room, and whether it is possible to cycle around an imagined room with obstacles without tilting. The comparisons assimilate the imaginings of physical thought experiments to the observation of physical experiments by using as a middle term for the

comparison mental processes which have one foot in the mind and one in the physical world, such as the non-conceptual and proto-conceptual contents typically discussed in philosophical theories of perception.

To bring out this point, consider Dretske's concept of simple or non-epistemic seeing: I may not have formed any beliefs about how many children were playing in the courtyard when I saw them, but I may be able to retrieve the information subsequently if I'm asked, by remembering the scene and picking out a plausible number of figures in it (Dretske 1981, 135-153). Dretske's theory exploits the fact that the information conveyed causally, 'naturally', during perception is dense and analog before being conceptually and propositionally encoded in the formation of beliefs. Gendler's theory of thought experiments exploits similar traits of mental imagery construed as a store of information about the world, only that Gendler additionally concentrates on the active, sensorimotor uses of such non-conceptual information: I possess information, in some quasi-natural form, about the natural world; retrieve that quasi-perceptual information; interact with it in a sensorimotor and affordance-based manner; and draw conclusions about the physical world on the basis of those interactions. So Gendler's answer to our question about how we can experiment with thoughts about physical process is that, at the relevant moments, 'thought experiments' are not *thought* experiments but a *quasi-physical* experiments.

To illustrate the possibility of acquiring such non-inferential knowledge, Gendler presents an experiment on mental imagery by Daniel Reisberg (1996). In Reisberg's experiment, subjects are (a) shown a form, (b) asked to memorize it, (c) to imagine it rotated, and (d) to draw a picture of the rotated form. When presented with a familiar geographical form rotated by 90 degrees, no subject succeeded in recognizing the form by rotating the image mentally (step c), but many were able to recognize it in their own drawing (step d). Gendler points out that these subjects have now acquired a 'new justified true belief (that the rotated image resembles Texas)' without 'inductive or deductive reasoning from known premises' (Gendler 2004, 1161). The following comparison is then made between the mental procedures involved in the Reisberg experiment and those involved in Stevinus's thought experiment:

> What's important for my purposes is the extent to which this case [Stevinus's] resembles those described above [Reisberg's]. Contemplation of an imaginary scenario (the cut string laid atop the prism) evokes certain quasi-sensory intuitions, and on the basis of these intuitions, we form a new belief about contingent features of the natural world (that the weight of four balls offsets the weight of three balls). This belief is produced not inferentially, but quasi-observationally: the presence of the mental image plays a crucial cognitive role in its formation. (Gendler 2004, 1161)

It is true that in both cases, Stevinus's and Reisberg's, belief-formation requires the presence and manipulation of a mental image. But as we saw in the analysis of Mach's concept of mimicry, this could be a trivial truth about mental representation, not a substantial claim about what determines the outcomes of thought experiments. In this latter respect, the Stevinus and Reisberg cases do *not* resemble each other. In Reisberg's experiments, the necessity of mental rotation for recognition is shown because of similarity between source-processes and target-processes: the experiment is *about* a mental process, visual recognition, and it is *conducted by* a mental process, rotation of mental images. Since the pictures preserve the features which are required for object-recognition procedures to operate, the relevant similarity constraint is met: it could be said that the picture-perceptions are simulations of object-perceptions. This contrasts starkly with Stevinus's thought experiment, which, while it is conducted mentally, is about a physical process whose relevant features it cannot replicate. Note what we *can* say about the Stevinus case: that the *mental visualization* of the chain on the prism resembles the *perception* of a chain on a prism in many relevant respects. But the respects in question are perceptual, and the thought experiment is not about the *perceptions* of chains on prisms, it is about the causal properties *internal* to systems comprising chains and prisms. The behaviour of those internal properties is *predicted* neither by process-driven simulation nor by contemplation and manipulation of the mental image, but by inductive application of prior implicit knowledge *to* the mental image as if it were not an image but an object.

How does this square with Gendler's conception of non-inferential knowledge? In the passage cited, the key claim lies in the concept of *quasi-observationally derived belief*. The concept is not only contrasted to inferentially derived belief, but has also to designate what *replaces* inference; and since there is no actual observation, it is the expression 'quasi' in 'quasi-observational' which bears the weight of the explanation. Consider first the following example of a 'quasi-*F*' which appears to be efficacious in the framework of a simulation. During our psychological responses to fictions, we sometimes experience emotions which are qualitatively similar to emotions that we experience in real contexts, but which lack the cognitive causes of real emotions, namely, beliefs. Quasi-emotions make fictions useful for simulating real emotions; and since emotions are among the springs of action, this in turn should make fictions useful for predicting actions in hypothetical contexts. Thus, in the expression 'quasi emotion', 'quasi' means two things. First, it means 'not a real *F* but similar to an *F* in its phenomenal quality'. Second, since the phenomenal quality of an emotion is a causally relevant property of an emotion, this implies that a process which can instantiate that property will bear a relevant similarity to the process which causes real actions. So by virtue of their similarity to real emotions, quasi-emotions confer on the experience of fictions the *capacity to simulate the processes by which action is determined*, and by virtue of this, *the capacity to predict action.*

Now, when we state that in thought experiments, beliefs about the physical workings of the natural world are *quasi-observationally* derived, do we mean that there is some mental process which, once put into action, can simulate the outcomes of physical processes by virtue of a relevant similarity to physical processes? No, because in the expression 'quasi-observational', 'quasi' applies to the similarity between visualization and perception, not to any similarity between on one hand physical processes, and on the other, the process by which *from* certain initial visualizations (inputs) we reach *further* visualizations (outputs). This contrasts with cases, mentioned by Gendler in support of her argument, in which individuals overcome their fear of flying by imagining flying safely (Gendler 2004, 1160). Gendler's point is that in such cases we form a new belief that fying is safe, not on the basis of deduction or induction, but just on the basis of imagining. These cases closely resemble the simuations just described involving quasi-

emotions, and they succeed because both the source- and target-processes are mental. But for that reason, they cannot solve the problem of physical thought experiments, which would have to be mental-physical simulations.

Gendler's examples bring into focus a further difference between thought experiments and simulations which should be a source of worry for the simulation hypothesis on thought experiments. The example in which we imaginatively fit elephants into a livingroom involves remembering a *particular* room and the spatial properties of members of a *particular* natural kind. This reminds us that simulation is well suited for rehearsing the behaviour of targets with *determinate spatial properties*, and therefore for dealing with contingencies to which physical laws are blind. This conflicts with physical thought experimentation, in which we tend instead to abstract from contigencies. The usefulness of fine-grained non-conceptual contents, proto-propositional mental contents, short-term memorization of the contents of simple perceptions for later retrieval, short-lived indexical concepts, and egocentric spatial representation, is context-relative. If we have an ability to mentally rehearse such events as getting the piano out of the window, it remains possible that that ability *as such* will not be suitable for extracting general conclusions about the physical world. The difference is worrying because it may mean that while we have an ability to reason spatially, and even to combine conceptual thought with perceptual or imagined contents, this ability may be exhausted by the production of *mental models* in physical thought experiments, without implying the presence of any form of process-driven simulation. An insufficient conceptual and cognitive analysis of thought experiments may lead us to confuse these processes of mental modelling with process-driven simulations. For example, 'simulation' may just mean that we inductively project prior knowledge, whether implicit or explicit, to perceptual-style imagery *as we would* to objects of perception.

This is borne out by work carried out on mental modelling. Mental modelling is a widespread activity found not only in thought experiments but in the navigation and manipulation of physical objects and in the formation, possession and revision of concepts. It seems to be the natural inheritor of Mach's project, but provides a clearer

position on what determines imagined outcomes during the mental manipulation of imagined objects. Instead of appealing to process-driven simulations, it appeals ultimately to constraints which are embedded, usually implicitly, in *concepts*. Concepts are described by Nersessian (2002) as sets of constraints on generating occurrent mental models; so on this theory, it is possible that what constrains the relation between imagined inputs and imagined outputs in thought experiments is concept-possession and mastery. The kinds of mental models used in physical thought experiments are defined as mental representations which preserve the salient spatio-temporal and causal relations of target objects (Nersessian 2002, 141, Johnson-Laird 1989). Thus, suppose that I attempt to predict the outcome of bending a rod. The rod-representation will have to be of something isomorphic to a rod, not to a spring or to a stone, in order for the knowledge implicit in the constraints on my concept for a rod (rather than my concept for a spring or a stone) to be triggered and applied. 'In order for' here introduces a causal relation between, on one hand, the structure of the mental representation, and on the other hand, the application of prior inductively formed knowledge. The operative processes are the isomorphic nature of the representation, and the inductively formed set of constraints on concepts. Thus, there is *no simulation of physical processes* occurring in the rod, but only a simulation in two other senses. (1) There is simulation *of a perception of* a physical state. This triggers (2), existing affordance-based knowledge about the outcome of the state, a process which involves no simulation. (3) There is inductive application of that knowledge *as if* we were applying it *in vivo*. For imagined outputs to be determined by a process-driven simulation, we would have to reach them by submitting the input-representation to a causal process which bears metaphysical similarity to the causal properties of the rod. Not only *do* we not do this, it seems that we *cannot* do it – we cannot submit a mental representation of the rod to a process of flexion. This is, of course, evident; what is not evident to start with, and what I hope it makes clear, is that what decides the outcomes of physical thought experiments cannot be process-driven simulations.

The situation is altogether different where mental-*mental* simulations are concerned, because these simulations comply to the relevant similarity constraint. Moreover, the power of simulation to individuate cases by replicating contingent local features (which are irrelevant for formulating physical generalizations) suggests that it is a good method for individuating action types and rehearsing practical reasoning. Practical reasoning typically does not draw conclusions from generalizations (in the form of *prima facie* premises) and has to be sensitive to local contingencies. Our mental-mental simulation abilities are thought to have evolved for the purpose of mindreading, which is the understanding of the epistemic states of individuals in given contexts. As such, mental-mental simulation also looks useful for rehearsing reasons and justifications in moral thought experiments, since these typically involve imaginatively placing subjects in concrete situations.

My main purpose in this section is to acquire the means to address the problem of *moral* thought experiments in the following section. Moral thought experiments are thought to acquaint us imaginatively with the epistemic situations of agents in moral dilemmas. So in this section, I will discuss the requirements for moral thought experiments so construed, by asking whether relevant similarity in fact obtains, and reviewing some problematic areas of mental-mental simulation which could be relevant to the moral cases. I will draw largely positive conclusions about the possibility and usefulness of mental-mental simulations. In the following section, however, I will argue against the usefulness of construing moral thought experiments as mental-mental simulations.

A key question about the processes required for mental simulations is whether, and how, we can simulate the causes of irrational behaviour, such as weakness of the will, using rational processes. Suppose that we adopt a simulation theory based on application of rational procedures and norms, such as that described in section 2. As it stand at least, such a simulation will lead to false predictions in any target area where non-rational processes play a causal role. When such breakdowns of rationality are failures in the

physical mechanisms *implementing* rational procedures, they cannot be predicted by using *mental* simulation. Suppose that I try to predict whether a smoker will resist buying his next packet of cigarettes. To make a valid probabilistic prediction, we should take into account processes which override rational procedures altogther, such as chemical reward pathways created by habit. So we cannot reach a valid prediction simply by placing ourselves in an imaginary situation *as agents*, but have also to apply a theoretical generalization. This problem cannot be avoided by construing the brain mechanism in terms of its conscious, phenomenal manifestation, namely, an urgent desire that we smoke. For *thus* construed, as a desire, the cause enters the realm of reasons and (under *ceteris paribus* conditions) we cannot but rationally reach the all-things-considered conclusion that the agent *prefers* to abstain. When norms of rationality are used to simulate, they tell us what an agent *ought* to do, not what he *will* do. To know what the agent *will* do, either the rationality-based simulation has to be completed with prior theoretical knowledge, or else some way needs to be found of simulating the causal role of states other than beliefs and desires.

A particularly important question in this respect is whether, and how, the causal role of *emotions* can be predicted by rationality-based simulations, because according to some authors (Tappolet 2003, de Sousa 1987), emotions can contribute to causing actions independently of reasons by focusing attention on aspects of the agent's situation that reasons fail to capture. For present purposes, two possible claims have to be distinguished concerning the emotions. One is that emotions have *rational underpinnings*, the other is that emotions *comply with rational constraints* when they contribute to determining actions. A striking version of the first claim can be found in the theory that emotions are perceptions of values (Mulligan 1998). According to this theory, emotions stand to values roughly as sensations stand to objective properties: they are the form in which the practical values and affordances of objects and situations are signalled to us in experience. For example, the emotion of fear is a perception of danger; the concept *danger* is a thick evaluative concept with a negative polarity, and, *ceteris paribus*, it implies disvalue. On such a theory, emotions are rational in the sense that they exist in the first place for reasons which comply with our interests (assuming an initial set of

priorities, or 'system-objectives' in Dennett's terminology). If such a theory could be generalized, it may show that evaluations are already built into the *descriptive* contents under which we represent imaginary situations, and if this is so, the imaginings should cause at least a *representation* of the appropriate emotions. Now, even this strong form of cognitivism about the rational underpinnings of emotions does not mean that a *given* emotion has to comply with rational constraints when it contributes to determining an action. For example, sentimental or sexual jealousy may have rational underpinnings under the set of priorities which initially set up our capacity to experience jealousy, so that jealousy may be said to exist in the first place to alert us to the presence of a certain kind of danger. All of this constitutes in an externalist sense a rational undepinning of jealousy, but it still does not mean that a particular episode of jealousy cannot be a cause of irrational action. The phenomenal character of jealousy, as distinct from its cognitive content which can enter propositional thought, can persist and cause a subject to act solely on its account, in a manner not too different from that in which the content of the smoker's desire can enter rational calculations, but not its motivational force.

However, the simulationist can appeal here to a body of evidence which suggests that the emotions can in fact be simulated, and to a concept of *imaginative acquaintance*: knowledge we acquire about situations which are described in sufficient detail from a subjective point of view, as generally occurs in literary or artistic fictions. For the concept of imaginative acquaintance to work, a plausible case has to be made that such emotions are indeed felt in the absence of beliefs, since the contexts of imaginative acquaintance are hypothetical. The evidence that such emotions are indeed engendered comes partly from the experience of fictions, which has led philosophers to formulate the 'paradox of fictional emotions'. Analysis of the premises of the paradox suggests both that these 'fictional emotions' are *qualitatively* similar to ordinary, belief-based emotions, and that they do not imply the presence of beliefs required for the emotions in non-fictional contexts (Radford 1975, Walton 1990, Ch. 7). There are also possible cognitive explanations of how the emotions in question could be engendered. Explananations include Gregory Currie's and Ian Ravesncroft's (2003) theory of belief- and desire-like imaginings; Shaun Nichols's (2004) single code hypothesis; to a lesser extent (due to its

less empirical and more phenomenological emphasis), Kendall Walton's (1990) claim that there is a special propositional attitude of 'make-belief'; and Damasio's (1991) theory that emotions are activated in hypothetical thought in order to guide practical reasoning. Gaut (2008) gives us an illustration of how imaginative acquaintance works when he holds that William Styron's novel, *Sophie's Choice*, acquaints the reader with the phenomenology of being in a situation he has not experienced himself. The experience of Styron's novel suggests that fictions can give us information without which we cannot account properly for either the *consequences* or the motivations of certain moral choices: the novel describes the progressive destruction of an individual by the emotions engendered by her own action, and we cannot make sense of how the emotions cause the moral destruction of the agent unless we have some notion of their phenomenal quality. (Further examples, and explanation of how theories mentioned here are applied to them, are given in section 5.)

To sum up, certain brain processes cannot be simulated mentally if our brain is not like the target's (for example, if it does not have the same reward pathways), forcing us to admit that theoretical knowledge is sometimes at least jointly required for predicting and understanding the behaviour of target-agents. However, there is plausible support for the thesis that mental simulation of emotions and their causal role is possible. With these proposals in hand, we can now examine the role of mental-mental simulation in moral thought experiments.

MORAL THOUGHT EXPERIMENTS

Now I will draw on the theories described in the preceding section to flesh out, and then to criticize, the thesis that mental simulation can explain the usefulness of moral thought experiments. The problem with the hypothesis, as I see it, is that it presupposes that useful *epistemic*, as opposed to subject-transcendent, discoveries can be made about the nature of values. This bias is explicit when thought experiments are intended to imaginatively acquaint us with the sentimental and emotional circumstances of moral

dilemmas. It is less obvious, but equally present, when thought experiments simulate the situation of a decision-making agent. Even simulation understood on the model which accounts least for the emotions, namely Dennett's normatively rational intentional stance, presupposes a set of hierarchical objectives proper to the system ('system objectives'), so that adoption of the intentional stance towards the target-system means thinking as if we had its own set of system objectives. Thus, even on a model which does not account for moral sentiments, the simulator will still represent the hypothetical situation *from the target-agent's point of view*. But accounting for moral values inevitably means having at some point to account for conflicts *between* sets of system-objectives, and this suggests that the point of view to be adopted should be *external* to the target-agent's. 'External' here means not so much 'allocentric' – since in simulating we adopt another agent's egocentric viewpoint, which means that we are being allocentric – as simply 'objective', or 'non-epistemic'.

Representation of a hypothetical situation from the target-agent's point of view is the common denominator of the concepts of 'belief-like imaginings' and 'desire-like imaginings' in Currie, pretense in Nichols, and make-believe versions of propositional attitudes in Walton: they are claimed to be sufficiently independent of our real beliefs and desires not to conflict with them, so that we can separately assume such attitudes while bracketing our own, putting ourselves in another, or a hypothetical, individual's place. (The same point is made in Leslie's tea-party experiment (Leslie 1994)). Simulating a decision-making situation adds a number of epistemic features which according to simulationists are absent when we merely represent it mentally. One feature is informational: we reproduce the individuality of the situation from the point of view of the putative target-agent immersed in it, therefore, we possess more information than we would as an external observer necessarily situated elsewhere. This additional information is represented mentally by means of imagined beliefs, perceptions, desires and preferences. A second, and crucial, feature is phenomenological: the beliefs, perceptions or desires should normally (for the simulation to comply with the simulation theory of mind) not be represented *de dicto* but *de re*: it is not a case of 'I imagine that X believes that *p*', but a case of 'I imagine (of myself) that I believe that *p*'. In perceptual imaginings

such as visualizations, the *re re* form can also have a simple formal object: 'I imagine of myself that I see *x*'. A third feature is acquaintance with the subjective quality of emotions as outlined in the previous section. It is distinct from the second feature because it concerns emotions as distinct from propositional attitudes, and because a simulation theory could include the first two features while excluding the third. These internal-viewpoint descriptions of moral dilemmas, involving hypothetical adoption of the agent's beliefs and preferences and potentially imaginative acquaintance with feelings, sentiments and emotions, may be contrasted with external-viewpoint descriptions of moral dilemmas, such as those utilitarians are supposed to use, which are not descriptions from *any* agent's point of view.

Such being the nature of the simulations in question, I turn now to my criticisms of the way they include epistemic features into moral thought experiments. If my criticisms are valid, then either moral thought experiments are valid because they are *not* simulations, or else they are invalid (unreliable for reaching theoretical propositions in ethics) because they *are* simulations.

The mental simulation hypothesis for moral thought experiments would say that process-driven simulations imaginatively acquaint us with epistemic states which are relevant to moral deliberation. This raises two problems: (1) *If* we accept that such epistemic information is indeed useful for knowing values, what guarantees that any epistemic information we do so obtain is correct? (2) What justifies the assumption that such epistemic information *is* indeed useful for knowing values? The problems are closely connected; for simplicity, I will not always distinguish them in what follows.

Consider the following case of prediction-failure, reported by Stich & Nichols (1992). Asked what a subject would do if it were asked to choose between two apparently similar objects, one to the subject's left, the other to the right, I will predict that they have equal chances of taking either. In fact however, *in vivo*, subjects display a greater propensity to choose the object to the right, so my prediction about the hypothetical situation fails. The first problem this case raises for the hypothesis is, how do we know that simulations of

moral dilemmas will not fail to accurately predict the processes in the subject which contribute to his choice, just as they do in the preceding case? This problem is aggravated by the fact that the kinds of philosophical propositions moral though experiments seek to formulate will not be verifiable independently (in the way the left-right choice is) and will not be very numerous, leaving us with hardly any scope for testing them.

The second, and related, problem is one that opposes rationalists and sentimentalists about moral value. Cases of prediction failure favour rationalists. When I attempt to simulate the left-right choice, rationality intrudes and overrides the unconscious processes which *would* make me choose the object to the right *in vivo*: since I am told that the objects are identical, I will *rationally* infer that the choices are equivalent, and (presupposing that the target subject is rational) will attribute to the target hypothetical subject a random choice. In real as opposed to hypothetical circumstances, the rational procedure will not override the unconscious processes which lead to the irrational decision. Nevertheless, it remains that *the rational procedure is the one that reaches the right prediction from a normative point of view, even if it is inaccurate from a descriptive point of view*. Agents should choose randomly, even if they are so constituted that they do not choose randomly. According to the imaginative acquaintance thesis, simulation plays the role of adding processes *other* than inference to decision-making; but if a mental simulation succeeded in just this respect, what would guarantee that the factors it introduced were not illegitimate for moral deliberation?

Consider the use of simulation and imaginative acquaintance in a familiar thought experiment by Bernard Williams. If we bracketed for the moment the detailed information given in the thought experiment's imaginary scenario, we would state the bare structure of the moral dilemma presented by Williams as follows: all things being equal, should we choose to kill one innocent individual even though we do not want to, or to let twenty innocent individuals be killed by someone else? For a consequentialist, there are more than one ways to reach the decision, but all the ways should lead to the former choice. Williams's purpose is to refute consequentialism by bringing out factors which are relevant to the moral decision but which cannot be accounted for by

consequentialism. To bring out these factors, he formulates the dilemma by using the following thought experiment:

> Jim finds himself in the central square of a small South American town. Tied up against the wall are a row of twenty Indians, most terrified, a few defiant, in front of them several armed men in uniform. A heavy man in a sweat-stained khaki shirt turns out to be the captain in charge and, after a good deal of questioning of Jim which establishes that he got there by accident while on a botanical expedition, explains that the Indians are a random group of the inhabitants who, after recent acts of protest against the government, are just about to be killed to remind other possible protestors of the advantages of not protesting. However, since Jim is an honoured visitor from another land, the captain is happy to offer him a guest's privilege of killing one of the Indians himself. If Jim accepts, then as a special mark of the occasion, the other Indians will be let off. Of course, if Jim refuses, then there is no special occasion, and Pedro here will do what he was about to do when Jim arrived, and kill them all. Jim, with some desperate recollection of schoolboy fiction, wonders whether if he got hold of a gun, he could hold the captain, Pedro and the rest of the soldiers to threat, but it is quite clear from the set-up that nothing of the sort is going to work: any attempt at that sort of thing will mean that all the Indians will be killed, and himself. The men against the wall, and the other villagers understand the situation, and are obviously begging him to accept. What should he do? (Williams, 170)

The factors Williams has in mind are 'moral feelings' of not being able to 'live with' what one has done, and the sense that 'each of us is specially responsible for what he does, rather than for what other people do' (Williams 173,171). By their very nature, feelings and the sense of selfhood cannot be appreciated without focusing some attention on the subjective states, and this is what Williams's literary fictional narration in intended to do, unlike my bare bones description further up. For Williams there should strictly speaking be no such bare bones description of the dilemma because, precisely, moral feelings and a sense of moral integrity ('each of us is specially responsible for what he

does, rather than for what other people do') *do* enter decision-making. The concept of integrity is cut out to distinguish the agent from other agents, so it implies situatedness, a subjective viewpoint, or egocentricity (in the technical, not the evaluative sense), which cannot be conveyed without the fictional narrative and its literary devices.

Moral feelings and damage to the sense of integrity are not mentioned or described in the fiction, but we are put in a position to appreciate them. A simulationist could hold that moral emotions can be rehearsed mentally in the form of quasi-emotions (described in section 3) or fictional emitions (section 4), and that the narrative causes pretend-beliefs (Nichols and Stich 1997, Nichols 2004) and pretend-desires (Currie 2002, Currie and Ravenscroft 2003). The sense of selfhood, which is required for the sense of moral integrity, could be conveyed by priviledging the agent's subjectivity over the subjectivity of the other individuals implicated in the dilemma. If the fiction is appreciated by causing *de re* thoughts, then the reader of the thought experiment is in fact placed imaginatively in the situation the dilemma describes. For example, the reader may imagine himself shooting a single peasant at the stake. This does in fact appear to be conveyed by the narration: some details are perceptual, and 'a heavy man in a sweat-stained khaki shirt *turns out to be* the captain in charge' is a description from Jim's viewpoint, not the narrator's. We are also given insight into the agent that we could only have by *introspection* ('Jim, with some desperate recollection of schoolboy fiction, wonders whether if he got hold of a gun, he could [...]'). In analyzing simulation theories of physical thought experiments, we saw that it is tempting to think there is some form of priviledged mental access to nature; in moral thought experiments, the special mental access is to other minds, and in particular to what it feels like to be in another's mind. In this case, though, the arsenal of theories described in Section 4 lends the thesis considerable credence: we do seem to have ways of knowing 'from the inside', though to a certain fallible degree, the mental life of real or hypothetical others.

Now, for Williams's argument to work, he has to show that the consequences (of killing one of the peasants) for the agent's moral integrity and moral feelings cannot be factored into a consequentialist calculation. Otherwise, the consequentialist will avail of them and

simply use the thought experiment to add an extra factor into his calculations. To prevent this, Williams argues:

> [...] we are partially at least not utilitarians, and cannot regard our moral feelings merely as objects of utilitarian value. Because our moral relation to the world is partly given by such feelings, and by a sense of what we can or cannot 'live with', to come to regard those feelings from a purely utilitarian point of view, that is to say, as happenings outside one's moral self, is to lose, in the most literal way, one's integrity. At this point utilitarianism alienates one from one's moral feelings; we shall see a little later how, more basically, it alienates one from one's actions as well. (Williams, 173)

Taken in isolation, part of Williams's argument is sentimentalist: since we have *moral* feelings which conflict with utilitarian moral judgments, we cannot be utilitarians. But this alone does not establish that negative moral feelings resulting from damage to moral integrity *cannot* be factored into consequentialist calculations. So Williams has to rely on further arguments, alluded to at the end of the citation: that the concept itself of an *agent* presupposes that of integrity. The arguments can be joined by using Williams's distinction between internal and external reasons. If Jim acts against his sense of moral integrity, he will be acting on external reasons (which are internal reasons to, among others, the nineteen peasants). That which allows the motivation of action and the concept of agency in Williams's perspective is partly something akin to Dennett's system-preferences which are *proper* to a system. The difference between 'selves' comes down to two factors: (a) one system's system-preferences (desires) concern *it*, not some other system; (b) we are constitutively (in terms of the phenomenology of agency, but also biologically) attached to *our* desires. So Williams's argument comes down to saying that we are not utilitarians because we have moral sentiments which conflict with utilitarianism, and because we could not be agents if we never acted on internal reasons.

Since my purpose is to examine the issues raised only insofar as their resolution can be affected by the use of thought experiments in ethics, I will restrict myself to mentioning

what I think are the relevant replies to Williams in that particular respect. In this respect, I will argue that Williams's position is weakened – not strengthened, as may be supposed – by the use of moral thought experiments involving mental simulations. Specifically, Williams's assumption that epistemic information internal to an agent is useful for knowing values can be challenged from two directions: with an ought/is argument, and with a methodological argument.

We saw in the case of prediction failure that the rational procedure is the one that reaches the right prediction from a *normative* point of view, even if it is inaccurate from a *descriptive* point of view. The problem with Williams's use of the thought experiment is that it priviledges the descriptive point of view: the mental simulation reproduces processes pertaining to the phenomenology of selfhood and agency, and to the generation of feelings. It is a descriptive device which shows us how we would feel under certain conditions *because we are so constituted,* in a sense not unlike that appealed to by moral sense theorists such as Hutcheson. So the consequentialist could argue that the kind of knowledge the thought experiment provides is knowledge about how we *are*, whereas the issue is to know how we *ought* to be, and what we should do. The consequentialist appears to have normativity on his side in this dispute, whereas Williams has only an appeal to the authority of description. In fact, the consequentialist can have a more complex position than the sentimentalist, for he can susbscribe to one theory concerning the way we are constituted, and to another concerning how we should try to become – *just as* we do in non-moral context, when we say that 'the way we are', which makes us choose objects to our right over objects to our left, is not the way we should be, even for our *own* good. On the possibility of becoming a consequentialist of *some* kind, consider the following sequel to Williams's story. Suppose that Jim makes the choice which means that he has to give up his sense of moral intactness and his emotional serenity. There is a clear sense in which his action is commendable not just in terms of the greater good, but as self-sacrifice *for* the greater good. But in virtue of what is his act commendable in the latter sense, if it is not because he opted for the way he should be, as opposed to the way he is? In defending the moral relevance of feelings and the sense of

selfhood, is the sentimentalist not asserting that *even if* a choice is somehow invalid, we should nevertheless stick to it, just because that's how we are?

The methodological argument says that apart from priviledging description, the thought experiment performs the function of priviledging the agent's point of view and is therefore *epistemologically egocentric*. Consequentialism presupposes precisely that this viewpoint be given up for an objective point of view so that all parties implicated by the dilemma are weighted equally in moral decision-making. In this sense, the thought experiment *qua* simulation or imaginative acquaintance hides the forest, the consequences for other agents, behind a single tree, the agent's subjectivity. In fact, the same thought experiment could also be used to show how an objective assessment can be clouded with epistemic illusions about the objective value of our personal integrity. It seems that to combat consequentialism, one would have to combat precisely its commitment to an objective viewpoint – and this, the mental simulation cannot do, because it operates from *within* an egocentric framework. To borrow terms used by Wilfrid Sellars in a different context, it 'operates within a framework, and cannot support that framework'.

Such arguments do not apply only to Williams's thought experiments, but suggest more generally that if moral thought experiments are mental simulations, then their use must priviledge certain moral theories – sentimentalist ones, for example – over others. Combined with the problem of prediction-failure, this should make them unreliable for formulating general propositions about the nature of value. If thought experiments are a reliable method for ethical thought, they must succeed in virtue of some procedure other than mental simulation.

REFERENCES

Brown, J.R., (2007) ‘Thought experiments’. Stanford Encyclopedia of Philosophy.

Carruthers & Smith (eds) (1996), *Theories of Theories of Mind*, 119-137. Cambridge.

Currie, G., (1996) ‘Simulation-Theory, Theory-Theory and the Evidence from Autism’. In Carruthers & Smith (eds), *Theories of Theories of Mind*, 242-256.

Currie, G. (1990) *The nature of fiction*. Harvard.

Currie, G. (2002) ‘Desire in imagination’. In Gendler & Hawthorne (eds) *Conceivability and Possibility*, 201-221. Oxford.

Currie, G. (1995) ‘Visual Imagery as the Simulation of Vision’. *Mind & Language* 10, 25-44.

Currie, G. & Ravenscroft, I. (2003) *Recreative minds*. Oxford.

Currie, G. and Ravenscroft, I. (1997) ‘Mental Simulation and Motor Imagery’. *Philosophy of Science* 64, 161-80.

Damasio, A. R., Tranel, D., & Damasio, H. (1991) ‘Somatic markers and the guidance of behavior: Theory and preliminary testing’. Levin et al., *Frontal Lobe Function and Dysfunction*, 217-229. New York.

Davidson, D. (1980) *Essays on actions and events*. New York.

Davies, M. ‘The mental simulation debate’. C. Peacocke (ed.), *Objectivity, Simulation and the Unity of Consciousness: Current Issues in the Philosophy of Mind* (Proceedings of the British Academy vol. 83), 99-127.

Davies, M. and Stone, T (1998) ‘Folk Psychology and Mental Simulation’. In O'Hear

(Ed.) *Contemporary Issues in the Philosophy of Mind*, 53-82. Cambridge.

Davies, M. and Stone, T. (2000) 'Simulation Theory'. Routledge Encyclopedia of Philosophy

Dennett, D. (1989) *The intentional stance*. Cambridge (Mass).

De Sousa, R. (1987) *The Rationality of Emotion*. Cambridge (Mass).

Dretske, Fred, (1981) *Knowledge and the flow of information* Cambridge (Mass).

Gaut, Berys (2008) 'L'apprentissage éthique, l'art et l'imagination'. *Ce que l'art nous apprend*, 23-35. Nancy.

Gendler, TS & Kovakovich, K. (2005) 'Genuine Rational Fictional Emotions'. *Contemporary Debates in Aesthetics*. Oxford.

Gendler, Tamar Szabo (2004) 'Thought experiments rethought – and reperceived'. *Philosophy of Science* 71, 1152–1163.

Goldman, A.I. (1992) 'In Defense of the Simulation Theory'. *Mind & Language* 7, 104-119.

Gooding, D., (1993) 'What is Experimental About Thought Experiments?'. D. Hull, M. Forbes, and K. Okruhlik (eds.) *PSA 1992*, vol. 2, 280-290. East Lansing,

Gordon, R.M. (1992) 'The Simulation Theory: Objections and Misconceptions. *Mind and Language* 7, 11-34.

Heal, J. (1998) 'Understanding Other Minds from the Inside'. O'Hear (Ed.) *Contemporary Issues in the Philosophy of Mind*, 83-99. Cambridge.

Johnson-Laird, P.N. (1989) ‘Mental models’. Posner, M. (Ed.) *Foundations of cognitive science*, 469-499. Cambridge (Mass).

Leslie, Alan (1994) ‘Pretending and believing: issues in the theory of ToMM’ *Cognition* 50, 211-238.

Mulligan, K. (1998) ‘From appropriate emotions to values’, *Monist: Secondary Qualities Generalised*, 161-88.

Needham, A. & Baillargeon, R. (1993) ‘Intuitions about support in 4.5-month-old infants’, *Cognition* 47, 121-148.

Nersessian, N. J. (2002) ‘The cognitive basis of model-based reasoning in science’. Carruthers, Stich & Siegal (eds.) *The Cognitive Basis of Science*, 133-153. Cambridge.

Nersessian, N. (1999) ‘Model-based reasoning in conceptual change’. Magnani, Nersessian & Thagard (eds.), *Model-Based Reasoning in Scientific Discovery*, 5-22. New York.

Nichols, S & Stich, S. (1997) ‘Cognitive Penetrability, Rationality and Restricted Simulation’, *Mind & Language* 12, 297-326.

Nichols, S. (2004) ‘Imagining and believing: the promise of a single code’, *Journal of Aesthetics and Art Criticism* 62/2.

Norton, J. (2004) ‘On Thought Experiments: Is There More to the Argument?’, *Philosophy of Science* 71, 1139-1151.

Norton, J. (2004b), ‘Why Thought Experiments Do Not Transcend Empiricism’. Hitchcock, C. (ed.), *Contemporary Debates in Philosophy of Science*, 44-66. Oxford.

Prinz, J. J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. Cambridge (Mass).

Radford C. (1975) 'How can we be moved by the fate of Anna Karenina?' *Proceedings of the Aristotelian Society Suppl. Vol.* 49.

Reisberg, Daniel (1996), 'The Non-ambiguity of Mental Images'. Cesare Cornoldi et al. (eds.), *Stretching the Imagination: Representation and Transformation in Mental Imagery*, 127-131. New York.

Slater, A., Morison, V., Somers, M., Mattock, A., Brown, E., & Taylor, D. (1990) 'Newborn and older infants' perception of partly occluded objects'. *Infant Behavior and Development* 13, 33-49.

Sorensen Roy (1992) *Thought experiments*. New York.

Stich, S. & Nichols, S. (1992) 'Folk Psychology: Simulation or Tacit Theory?' *Mind & Language* 7/1, 35-71.

Stone, T. & Davies, M. (1996) 'The Mental Simulation Debate: A Progress Report'. Carruthers and Smith (eds) *Theories of Theories of Mind*, 119-37. Cambridge.

Tappolet, C. (2003) 'Emotions and the Intelligibility of Akratic Action'. Stroud and Tappolet (Eds.), *Weakness of Will and Practical Irrationality*, 97-120. Oxford.

Thompson, J. (1971) 'A Defense of Abortion', *Philosophy and Public Affairs* 1/1, 47-66.

Walton, K. (1990) *Mimesis as make-believe*. Cambridge (Mass).