

René van Woudenberg

## Belief is Involuntary: The Evidence from Thought Experiments and Empirical Psychology

### *Abstract*

This paper argues for the claim that belief is involuntary. Evidence in favour of it comes from various thought experiments. However, other thought experiments might be taken to indicate that belief is not involuntary (thought experiments regarding such policies as the policy to consider only evidence in favour of a claim and to neglect contrary evidence, or the policy to join a group of believers in a claim, or the policy to apply some form of self-suggestion). It is argued that none of these thought experiments should lead one to reject the main claim of this paper. Some evidence from empirical psychology, viz. the evidence in favour of the phenomenon of “motivated reasoning”, may be thought to undermine some of my arguments in favour of the claim that belief is involuntary. It is argued that this is not the case. Some other type of evidence from psychology, viz. the evidence related to the unconscious, by contrast, may be thought to support the claim that belief is involuntary. It is argued that this is only partially so.

No one has a choice as to what to believe. No one can, simply by an exercise of his will, choose or decide to believe proposition *p*. Belief is involuntary, as I shall say. By way of clarification, suppose I now have no beliefs about where the world’s oceans are deepest and want to have that situation rectified. Then I can exercise my will in such a way that I wind up believing that the Mariana Trench, Western Pacific Ocean, is the deepest point in the world. All I have to do for this is to consult an encyclopaedia, or an authoritative source on the deep seas. The thesis that belief is involuntary does not mean to deny this. What it does mean to deny is that I can believe that the Mariana Trench is the deepest point in the world, just by deciding to do so. More generally, what it means to deny is that for any specific proposition *p* that we care to think of, we can bring it about, simply by exercising our will, simply by deciding to do so, that we believe it. As I will be thinking of it, the concept of “voluntary belief” (or “belief at will”, or “belief by fiat”) cannot be applied to a belief unless at least the following conditions are satisfied:

- (A) The belief is acquired independently of any truth-relevant considerations on the part of the subject.<sup>1</sup>
- (B) The subject is fully aware of the fact that he attempted to acquire the belief independently of any truth-relevant considerations.<sup>2</sup>

This paper is in support of the claim that belief is indeed involuntary, and by implication that no belief satisfies conditions (A) and (B). Section 1 offers a number of simple empirical arguments, thought experiments really, for that claim, and furthermore discusses three ways in which voluntary belief, notwithstanding those arguments, might be thought possible, viz. paying selective attention to evidence related to *p*, seeking the company of *p* believers, and some form of self-suggestion. It is argued that none of these work. Section 2 discusses some empirical psychological findings in order to see whether they support or controvert the claim that belief is involuntary.

If my conclusion is correct, if belief is indeed involuntary, this puts pressure on an assumption that many people, philosophers and non-philosophers alike, seem to commit themselves to, viz. that humans can be rightly held responsible for at least some of their beliefs.<sup>3</sup> This paper doesn't estimate the force of the pressure. That has to wait another occasion.

### 1. *The Evidence from Thought Experiments*

The arguments for the involuntariness of human belief that I am now going to discuss are "simple empirical" arguments because they don't make use of the elaborate empirical research as done at, for instance, psychology labs. The arguments are, in the main, based on thought experiments. In or-

<sup>1</sup> This condition is meant to rule out that cases like the following count as cases of voluntary belief: someone wants to believe, for some reason or other, that the door is open, and in order to bring it about that she so believes, she (first) opens the door and subsequently forms the belief that the door is open. In cases like these the belief is *not* acquired "independently of any truth-relevant considerations". Cases like these are due to Feldman (2001, pp. 81-82).

<sup>2</sup> Winters (1979, pp. 244-245) provides a somewhat similar account of "voluntary belief" (or as she says "believing at will"), but also specifies the condition "that the belief has been acquired directly and as a result of intending to hold it". As will become clear in the first section, I see no reason to accept this condition. A belief (if such there is) that has been acquired *indirectly* (and as a result of intending to hold it), might still be a case of "voluntary belief".

<sup>3</sup> Evidence for the claim that many people, as well as many social institutions, are committed to the idea that humans can rightly be held responsible for their beliefs is presented and discussed in Woudenberg (2009a). Furthermore, an analysis of the practice of excusing ourselves and others for certain bad actions because those actions were performed out of non-culpably ignorance, also indicates how deeply ingrained the commitment is to holding each other responsible for what we believe. See for this Woudenberg (2009b) and Peels (2012, ch. 5).

der to see what these arguments amount to, we must first note that the complement of involuntariness is a rather rich set. Something can be voluntary in more than one way, and in order to refer to those ways I am going to use distinctions and labels due to William Alston (1989, p. 137; 2005, pp. 58-80). To begin with, something can be voluntary for a person in the way that, for a healthy person, raising his arm is voluntary. All that is needed for that is to carry out the intention to raise one's arm in one uninterrupted intentional act. A healthy person can raise his arm "just like that", without having to return to the attempt a number of times. If someone *cannot* raise his arm "just like that" and has to make more than one attempt (perhaps because he suffers from a muscular disease), that act is not voluntary in the sense now intended. This kind of voluntariness might be labelled *basic immediate control*. The reason why it is called "immediate" has to do with the "just like that" character of acts that are voluntary in this sense.<sup>4</sup> The reason why it is called "basic" will become clear upon considering a second way in which an act can be voluntary.

Consider such actions as starting one's car's engine and making someone happy by paying him a compliment. One can do these things, just as one can raise one's arm, in one uninterrupted intentional act, "just like that", in one try. But in order to succeed in doing these things more is needed than just a volition on the part of the agent. In each of these cases the agent's bodily movements (movements of the hand, and movements of the vocal chords, respectively) must have certain consequences – consequences that are either causally (in the case of starting the car) or conventionally (in the case of making someone happy) connected with his bodily movements. Starting the car's engine and making someone happy (in the way indicated) are certainly voluntary acts – and, if circumstances cooperate, they can be done "just like that". So these acts are under one's immediate control. But they are not 'basic' in the way that raising one's arm is, because these actions are the causal or conventional *consequences* of acts that *are* within one's basic immediate control. The kind of voluntariness involved in cases like these Alston labels *non-basic immediate control*.

There is a compelling argument for the conclusion that for humans in the actual world belief is neither under basic immediate voluntary control, nor under non-basic immediate voluntary control. The argument, offered by Alston (1988, p. 122), consists in asking yourself the simple question: can you, at this very moment, bring it about that you believe that the earth has three moons, just by deciding to do so? Can you now believe that the

<sup>4</sup> It should be noted, as Alston does, that these actions are immediate in a loose sense. In a strict sense only actions like forming an intention are immediate.

canals on Mars once held water, “just like that”, by *fiat*? In case you are insufficiently motivated to even try, suppose you are offered 500 dollars if you would succeed. Could you then do what it takes to get the reward? All of this seems utterly impossible. The conclusion is that belief is involuntary in the sense that we have no immediate voluntary control over it, neither basic nor non-basic.<sup>5</sup>

It is perhaps useful to consider a couple of cases that might be thought to controvert this conclusion. First there are cases of decision making under uncertainty. An officer in war time, a mountaineer who lost his way, parents whose child escaped from their eyesight have to decide on a certain course of action. They have to decide whether or not to launch an attack, which way to go, and where to look. It surely isn't the case that any decision here would be as good as any other – but still there are, or so we may suppose, genuine options. Whichever option is selected, however, one doesn't thereby *decide to believe* that that option will lead to success. In such cases one decides to proceed in a certain way; one does not *decide to believe* that proceeding in that way will bring victory, lead back to the right track, or guarantee that the child will be found.

Second, scientists may perform experiments on the basis of a certain assumption or working hypothesis just to see what results emerge. Depending on the results, such assumptions and hypotheses can be readily abandoned. Adopting and abandoning an assumption, however, does *not* mean that one first has decided to believe it and then decides to no longer believe it. Surely, in a case like this decisions are made. But what is decided is, first, to proceed on a certain assumption and next to no longer do that. Belief is kept at bay.

The field of the voluntary is not exhausted by immediate voluntary control, however. Consider the kind of voluntariness involved in losing weight through keeping a diet. Losing weight is not an action that can be performed “just like that” but requires repeated efforts stretched out over longer periods of time. We have voluntary control over the action of losing weight, albeit only in an indirect way. A further point here is that there exist considerable differences in the “strength” of the sort of control that we have over those things that are within the span of that sort of control. I may have a much firmer control over my weight than you have over yours. This kind of control, which comes in degrees, is not immediate; but it is indirect, as

<sup>5</sup> This conclusion also holds for propositions that are false. Moreover, the same sort of argument can be used to show that not only belief, but also disbelief and withholding are involuntary.

it is stretched out over a period of time. Alston calls it *long-range voluntary control*.

As indicated, there are differences in strength of long-range voluntary control. If that sort of control reliably brings about the intended effect, it is strong. But if it only occasionally brings about the intended effect, it isn't really control. For control requires that one can reliably, so in a high percentage of cases, bring about the intended effect. This holds for immediate as well as long-range voluntary control. For the discussion that is now to unfold, this will prove to be a rather important thought.

Is belief under long-range voluntary control? If it is, there will have to be things that we can *do*, and do them repeatedly, such that doing them will after some time reliably lead to the formation of the intended belief. (The belief involved will still have to be the belief that *p*, with *p* being a specific proposition.) What sort of things? The most plausible candidates might, perhaps, be the following:<sup>6</sup>

- paying selective attention to the evidence – only focusing on evidence in favour of *p* and turning one's head away from any evidence against *p*,
- seeking the company of believers in *p*, and avoid sceptics and unbelievers,
- apply some method of self-suggestion, such as, perhaps, repeating to one-self "yes, *p*!" or "I believe that *p*".

Let us look at these candidates somewhat more deeply, and suppose we start with the first: paying selective attention to evidence. A number of cases, or types of cases, have been described with the aim of establishing that paying selective attention to evidence gives us long-range voluntary control over at least some of our beliefs. One case, due to Roy Edgley (1969, pp. 63-64), features a man who is very sick and who is persuaded by his friends that it would be good for him if he were to believe that he will recover. The sick man is persuaded by them and accordingly tries to believe that he will recover by dwelling on favourable evidence, and by making specific plans about what to do after regaining health. Edgley believes this man may succeed. Is he right?

The question to ask is whether paying selective attention to the evidence relevant to *p*, can be used as a policy to insure that one comes to believe *p*. The question is not whether we do sometimes pay selective attention to evi-

<sup>6</sup> Note that, as is implied by conditions (A) and (B), the thesis that belief is involuntary is the thesis that belief *acquisition* is involuntary. It is a separate topic whether belief *maintenance* is involuntary. For discussion of the possibility of voluntary belief maintenance, see Wolterstorff (2010).

dence. “It is as easy to close the eyes of the mind, as those of the body”, said Bishop Butler (1889, p. 480), and virtually everybody who has pronounced on this topic says the same. There seems to be ample reason for that consensus, whether we like it or not. The question is not whether we sometimes pay selective attention to the evidence (we do) but whether paying selective attention to the evidence relevant to  $p$  can be adopted as a policy when we want to bring it about that we acquire the belief that  $p$ . Can we, by adopting such a policy, exert long-range voluntary control over what we believe?

I am very sceptical about this. If paying selective attention to evidence is to be a policy that can be adopted in order to attain the specific goal of acquiring the belief that  $p$ , then adoption of that policy must ensure, or make it highly likely, that the goal will be reached. For if it doesn't, adopting the policy doesn't provide control. (I am hence assuming that control over  $X$  requires the ability to reliably, i.e. with high likelihood, bring about  $X$ .) Now adoption of the selective attention policy, it would seem, *doesn't* make it highly likely, that the belief-goal will be reached. The point is that paying selective attention to evidence relevant to  $p$  so as to ensure belief in  $p$ , isn't something one can consciously *do*. My argument is of the simple sort used earlier on: try it – you won't succeed.

Applied to Edgley's case of the sick man, this works out as follows. As the description of the case already indicates, the hoped-for belief may *or may not* result from following the advice offered. There is no guarantee, not even some modicum of likelihood, that it will, as paying selective attention to the evidence relevant to  $p$ , doesn't reliably bring about the belief that  $p$ . And then there is the further point that it is very dubious that one can consciously adopt paying selective attention as a policy.

The point I am making is, to repeat, not that we never pay selective attention to evidence; the point is that we cannot adopt “paying selective attention to evidence” as a *policy* that gives long-range voluntary control over believing  $p$ .

Another thing one could think may secure that one acquires the belief that  $p$  is to seek the company of  $p$ -believers and avoid the company of  $p$ -sceptics and  $p$ -disbelievers. It is a well-known fact that meeting with a company of like-minded persons can be very reassuring and heartening. But this is not relevant to the discussion, for this company consists of persons who already share beliefs, whereas our discussion concerns whether someone can *acquire* the belief that  $p$  by joining a company of  $p$ -believers and shunning the company of non- $p$ -believers. Nor is the following very real phenomenon relevant to the discussion: people sometimes wish, for reasons of status, to become members of a certain social group, and, once entered, acquire certain beliefs that are popular among its members. This is not rele-

vant, for here the reason to join the group is not the wish to believe that *p*, but the wish to have one's status boosted. The sort of situation that *is* relevant to the discussion would be something like this: *S* wants to believe that *p*, and notices that the members of group *G* all believe that *p*; *S* then decides to join *G*, to hang out with its members, to do the things they do; in that manner, *S* thinks, he can reliably bring it about that he acquires the belief that *p*. If what *S* thinks is correct, he would have long-range voluntary control over the belief that *p*; but if it is not, then he would not.

One might think that whether what *S* thinks (viz. that by joining *G* he will ensure that he will acquire the belief that *p*) is correct, depends on what goes on in *G*. If its members, whenever they meet, say to each other again and again such things as “*p*”, and “yes, *p*!”, “hurray for *p*”, and “three cheers for *p*” (thus expressing their belief that *p*), but never really discuss *p*, never consider any evidence for or against *p* (or if there is no evidence, never discuss why the absence of evidence should be no problem), never relate *p* to social practices that might be based upon them, then it seems rather unlikely that *S* will begin to believe *p* *due to joining G*. Joining *G* can't be used as a policy to bring about the formation of belief in *p*. But suppose the members of *G* go about differently, and *do* discuss evidence relevant to *p* (and *do* conclude each time that *p*), etc, then, yes, *S* might wind up believing *p*, depending on the details of the discussion. If the discussion is biased, if relevant counter evidence is suppressed, if counter positions are “refuted” only after having been mischievously misrepresented, etc, *and if S has some awareness of this*, then again it is surely not to be expected that *S* will wind up believing what he wanted to believe *due to joining G*. But what if the discussion among the *p*-believers *is* fair, open-minded, and due attention is given to counter evidence, and contrary positions are criticized only after having been given a fair hearing, and *S has some awareness of this* – what then? Can *S* be assured, prior to joining *G*, that he will wind up believing what he wants to believe? Well, perhaps so, perhaps not, that is hard to tell in advance: it all depends on whether *S* will find the transactions of the *p*-believers compelling; perhaps the contrary positions have been described so fair that *S* is unmoved by the criticisms offered, or perhaps the social practices that are based on *p* (or involve *p* in some way or other) are so repelling, that *S* no longer finds belief in *p* attractive. But now suppose that *S*, prior to joining *G*, knows that many others that have joined *G* in the past, have wound up believing that *p* due to arguments and evidence relevant to *p*. Wouldn't that assure him that the same will happen to him and hence that his joining *G* will ensure that he will eventually believe what he wants to believe? Well, it seems likely that this will so ensure him. But the thing to see is that the case as described is irrelevant to the discussion. For if *S* joins *G* and comes to believe what he wants to believe, then

he doesn't come to believe  $p$  *by virtue of joining*  $G$ , but by virtue of something rather different, viz. *by virtue of the arguments offered by the members of*  $G$ , as well as *by virtue of the positively evaluated practices that involve belief that*  $p$ . And thus the case is irrelevant to the discussion, for then it is not a case of voluntary belief, as condition (A) for voluntary belief is not satisfied; the belief is no longer formed independently of truth-relevant considerations.

The conclusion should therefore be that whatever effects the joining of a certain company may have, *just* joining a company of  $p$ -believers cannot be adopted as a policy to ensure that one winds up believing what one wants to believe. And hence just joining a company of  $p$ -believers is not a way to have long-range voluntary control over the belief that  $p$ . (None of this implies a denial of the fact that one's joining a particular group might result in picking up beliefs that are indigenous in it.)

Let us now finally turn to the third and final candidate for long-range voluntary control over beliefs: self-suggestion. It is part and parcel of folk wisdom that when people say something long enough to themselves, they are bound to believe that it is true. And this seems to be the idea about what is going on in self-suggestion: you say to yourself " $p$ ", and "yes,  $p$ ", and " $p$  is true", "that famous Mr.  $X$  believed  $p$  – and he was no fool", and the like, and as a result you wind up believing that  $p$ . Self-suggestion, then, can be taken to be this: you suggest  $p$  to yourself by saying to yourself " $p$ " and the like over a stretch of time, and as a result you acquire the belief that  $p$ . When stated thus bluntly, it would seem that if self-suggestion is to be a policy one can consciously adopt in order to acquire the belief that  $p$ , it is severely limited in its application. It seems incredible that simply by repeating to oneself "the earth has three moons", one can bring it about that one starts believing so. And something similar holds for all sorts of factual propositions about history, geography, cosmology, culture, politics, jurisprudence and so forth. There would seem to be only two classes of propositions where self-suggestion might have a chance of success: propositions about oneself, and moral propositions. Let us consider these in this order.

The first class of propositions include: *I am a fine fellow; I have nothing to write home about; I can't do math; I can do the triathlon; I am not going to be nervous when I have to perform in front of the queen; I will get that job; when I am there I am going to make a fool of myself; there was nothing wrong in what I did; I am not going to feel embarrassed when I see her.* Can we, by saying (or thinking) these things to ourselves, reliably bring it about that we will eventually believe any of this?

This is a very difficult issue, but still some relatively uncontroversial points can be made. First, saying something to oneself about oneself, does seem to have genuine effects on the one saying (and hearing) it. The Frenchman

Emile Coué (1857-1926) taught his patients to say to themselves every morning and every evening: “Tous les jours, a tous points de vue, je vais mieux et mieux” (Every day in every way I am getting better and better). And he urged that these words be spoken unthinkingly: “don’t think what you are saying, say it as you say the litany at church!” This simple advice proved extraordinarily effective: the Coué movement swept across Europe like a prairie fire.<sup>7</sup>

Second, the genuine effects of saying something to yourself, however, may not be belief. Saying to yourself ‘I can do it’ may help you to actually do the triathlon. But even if it so helps you, that doesn’t mean that before crossing the finish you *believed* you could. It might be that you said to yourself “I can do it”, but all the while had the gravest doubts about whether you really could, in which case you didn’t have the *belief* that you could do it. But even then, saying to yourself that you can do it, can actually help you doing it. Isn’t this very curious? Well, it is as curious as the effects that exhortations *of others* can have on our performances. There is power in hearing someone else yelling to you “Come on!”, “You can do it!”, “Go on!”. In the case of self-suggestion the only difference is that the exhortation (the yelling) comes from yourself. So my point is that saying repeatedly to yourself that p may be effective in the sense that it may help you succeed in something you are trying to accomplish – even if it doesn’t make you believe that p.

Third, saying something to yourself may have a genuine effect other than that of an exhortation. When you repeat to yourself “I am not going to be nervous when I have to perform in front of the queen” while you are preparing for just such a performance, that may help you in the sense that when you envisage that prospective situation, so when you try to imagine what it will be like to perform in front of the queen, you get a feel for its difficulties and challenges, and thus you go there forewarned and hence forearmed. So, your self-addressed and oft-repeated speech has a genuine effect – but it isn’t the effect that you come to *believe* that you are not going to be nervous when you have to perform in front of the queen.

There is yet another genuine effect of repeatedly saying “p” to yourself: repeatedly saying to yourself “I am going to get that job”, might help you to stay focussed on a goal that you have set for yourself. It might keep your vision alive. But again, this self-addressed speech need not induce in you the belief that you are going to get that job.

<sup>7</sup> My source here is Weatherhead (1951, pp. 122-128).

So far I have only suggested that self-addressed *p*-speeches may have effects *other* than the formation of the belief that *p*. The crucial issue, however, is whether making such speeches reliably brings about the formation of the belief that *p*. The issue is, in other words, whether we can adopt addressing such speeches to ourselves as a policy in order to secure that we come to believe that *p*. If we can, we have long-range voluntary control over belief in certain propositions about ourselves. I can only report that it seems to me that we have no such control. I have never seen a convincing case of someone doing that; and all the cases that seemed somewhat plausible, are much more naturally understood as cases where the self-addressed speeches are exhortations, helps to be forewarned, or means to stay focussed. And what the previous discussion illustrates is that my denial needn't commit me to denying the genuine effects of self-addressed speeches.

Let us now move to the other class of beliefs over which we, perhaps, have long-range voluntary control: beliefs in moral propositions. One special problems needs to be set aside first, though. According to emotivists there are no moral truths. On that position, what it means to say "I believe that I ought to help my neighbour" is *not* that I believe *that it is true* that I ought to help my neighbour. It means something like "Hurray for helping my neighbour" – and thus expresses an emotion or other affective state on the part of the speaker. Since the emotivist's understanding of belief is radically different from what I take belief to be (*viz.* that to believe that *p* *is* to believe that *p* is true), and furthermore inherently very implausible, I shall pass it by.

So suppose we think of moral propositions in a realist sense, and suppose furthermore that someone wants to believe that *p*, where *p* is a specific moral proposition – can he then reliably bring it about that eventually he winds up believing that *p* simply by saying to himself "p", "yes, p!" and the like? I must confess to not being able to see that this can be done. When I repeat to myself that cheating is good as long as you aren't found out, I don't eventually wind up believing it at all. As a matter of fact, after having tried, it seemed to me as obviously false as it ever did. But what about moral propositions whose truth value is not clear to me? What about the proposition that it is my duty to help my neighbour's son with math, or the proposition that Lizzy is dependable, or the proposition that what Friedrich Weinreb did during world war II, was good – propositions about which I have my suspicions? Can I eventually bring myself to believing any of these, simply by repeating them to myself (so not by considering their truth value)? Can I, with respect to those propositions, adopt the policy of self-suggestion as explained, and thereby ensure that I will wind up believing them? This strains credulity. We simply have no such powers.

The upshot of this section, then, is that there are relatively simple empirical arguments, thought experiments really, for the conclusion that belief is involuntary.

## 2. *The Evidence from Empirical Psychology*

The topic of the previous section currently enjoys great interest from the side of empirical psychologists. In this section I discuss two lines of psychological research that are pertinent to the claim that belief is involuntary. The first line of research concerns so-called “motivated reasoning”, the second unconscious belief. My goal in doing this is to see whether what I have been arguing so far on the basis of simple empirical arguments finds any confirmation or disconfirmation in empirical psychological research.

The first line of research, on “motivated reasoning”,<sup>8</sup> is especially relevant to my claim that “paying selective attention to the evidence”, “seeking the company of p-believers” and “self-suggestion” aren’t policies one can adopt in order to ensure that one winds up believing p (if it is p one wants to believe). Let me first state what I take to be the fundamental conclusion of the research and next provide some illustrations of how this conclusion is reached.

The fundamental conclusion of the research, to a first approximation, is that motivation affects reasoning. But what is “motivation” supposed to be here, and what “reasoning”? To start with the latter: “reasoning” is not what philosophers typically mean by this term; it does not, or not exclusively, refer to the process that consists in the thought progression that moves from premise(s) to conclusion(s). “Reasoning” is an umbrella term for “the cognitive processes we engage in to arrive at a judgment” (Kunda 1999, p. 211). The literature I am referring to makes no systematic distinction between “belief” and “judgment”, and the terms are used more or less as synonyms. So in order to obtain at least a verbal connection with the topic of the previous section, we might say that “reasoning” refers to any process that results in the formation of belief. Such processes include: “forming impressions, determining one’s beliefs and attitudes, evaluating evidence, and making decisions” (Kunda 1990, p. 480), as well as modes of dealing with information. As to “motivation”: subjects in whom processes of belief formation are at work (subjects that engage in reasoning) will as a rule be driven by motivations – and the processes operative in the formation of belief (their reasoning) will be driven by those motivations. Or, as Ziva Kunda has it, their

<sup>8</sup> I am drawing here primarily on Kunda (1990, 1999).

reasoning will be motivated. Now there are motives and motives. Subjects can be driven by the motivation to arrive at an accurate conclusion, but they can also be driven by a motivation to arrive at a particular, as Kunda calls it, “directional” conclusion. Given these clarifications, the fundamental conclusion of the research I am referring to can be stated as follows: both the motivation to be accurate and the motivation to arrive at a particular conclusion affect the processes operative in belief formation (both motivations affect reasoning).

For present purposes only one half of this conclusion is relevant. The present purpose is, after all, to connect the research on motivated reasoning with the claim made in the previous section that “paying selective attention to the evidence” and the like aren’t policies one can adopt in order to ensure that one winds up believing  $p$  (if it is  $p$  that one wants to believe). And to that purpose it is only relevant that the motivation to arrive at a particular conclusion can affect belief formation.

How is this fundamental conclusion reached? By empirically finding out that cases like the following are not at all exceptional. The case is of a woman who throughout her three pregnancies continued smoking. It was the days when it wasn’t common knowledge that that might not be a good idea. When her children are grown up, she reads a newspaper article on the effects of smoking during pregnancy on babies. She exclaims to her daughter, who had read the article too: “This article is full of nonsense. It says that if you smoke during pregnancy you will have small babies. Well, my babies were huge!” And they were huge indeed. The daughter then points out to her mother that the article also said that smoking during pregnancy could cause lung problems in babies, and reminds her of the fact that both of her huge brothers had suffered from spastic bronchitis throughout their childhood. To which her mother’s dismissive response was “that is only a sample of two”.<sup>9</sup>

This, it is said, is an example of motivated reasoning. The reasoning here is motivated by the wish to arrive at a particular conclusion, viz. that the article is full of nonsense, and hence need not be taken seriously. This, it is said, is a case where someone comes to believe what she wants to believe *because* she wants to believe it. Other cases that have been adduced on a more hum drum level include: lovers that are blind to the blemishes in their loved ones, and parents that are blind to obvious faults in their children, etc. A more dramatic case is discussed by Randy Shilts in his book *And the Band Played On*<sup>10</sup> in which he describes the reactions of various communi-

<sup>9</sup> The story is related in Kunda (1999, p. 212).

<sup>10</sup> Shilts (1987). This case is also described by Kunda (1999, pp. 223-224).

ties to the ever more disturbing findings about the, at the time, unknown disease among gay men, now known as AIDS. The gay community, the scientific community, the U.S. government, and the blood banks, all initially refused to believe evidence for the devastating conclusion that the disease was infectious and could be transmitted sexually. Each group had its own motivation for believing that this horrific scenario was false. For the gay community a characteristic lifestyle was at stake, the blood banks would have to acknowledge that their blood was unsafe and to modify their procedures at great cost, the government would have to face vast financial and social implications, whereas the scientific community would have to openly declare that many lives, both inside and outside the gay community, were in jeopardy. And all of these groups were to some extent successful. Scientists followed every lead that pointed to alternative explanations, and within the gay community an alternative theory that postulated a right-wing conspiracy to undermine the gay lifestyle became popular.

If these cases are illustrative for directional motivated reasoning, then it is clear that motivated reasoning is a bias that pollutes the beliefs it engenders.

It is important to note that the empirical research also suggests that the power of motivated reasoning is limited. We are not at liberty to believe whatever we want to believe simply because we want to. Research indicates that the desired conclusions are only drawn when subjects can come up with enough evidence to support it, when they can construct a justification for the desired conclusion. In the case of the heavy smoking woman, the justification (evidence) for resisting the belief that her sons' lung problems were caused by her smoking habits, was that two children whose mother smoked during pregnancies suffer from lung problems, doesn't prove that the smoking *caused* the problems. In the AIDS case, the conclusion that the disease is infectious and sexually transmitted could be resisted so long as there seemed to be reasonable alternatives. But as the evidence mounted, these alternatives no longer seemed reasonable and virtually everybody came to believe the fearsome conclusion.

An important question now arises. In the cases mentioned (and the literature suggests that such cases, in all sorts of variations, abound) bias is at work. But why do psychologists think that the best account of these biases is in terms of "motivation"? Couldn't there be an account of the phenomenon (*viz.* resisting a certain conclusion) purely in cognitive terms – in which case it would be unwarranted to speak of a bias? To return to the AIDS case: couldn't it be said that, prior to the piling up of evidence that pointed to the fearsome conclusion, the evidential situation was simply unclear, and that various alternative hypotheses were live options, and that there was simply no compelling reason for thinking that the disease was infectious and

sexually transmittable? This account of the case is not in terms of “motivation” but purely in “cognitive” terms such as: evidence, hypothesis and reason. And doesn’t this account work as well as does the account in terms of motivation? If so, it would seem to be preferable because it does not invoke the extra notion of “motivation” – and we should always favour the more parsimonious account.

Psychologists claim, however, that there is independent evidence for the thesis that motivation affects belief (Kunda 1990, 1999, pp. 215-223) Various studies indicate that humans are strongly motivated to believe that other persons, about whom they know next to nothing but on whom they are going to be dependent, are likable – much more likable than when they were *not* going to be dependent on them. Someone who is about to go on a blind date with Mr X, will be strongly motivated to believe that Mr X is likable. And if you know that the success of your team depends on the competence of a new but unknown team mate, you may be strongly motivated to believe that this person is competent. That is the idea. The evidence for it is of the following sort. In one study (Berscheid et al. 1976)<sup>11</sup> participants take part in a dating experiment. Before meeting their dating partners, the participants watched a videotaped discussion among three people, one of whom was said to be their future partner. There were three groups of participants, each expecting to date a different one of the three discussants, and an additional control group that was not expecting to date any of them. The participants rated their expected date as more likable than the two other discussants, and as having more positive personality.

As Kunda says, results such as these are hard to account for in purely cognitive terms. After all, the participants knew (almost) nothing about their future date, and hence couldn’t have formed any specific expectations about him or her. Kunda’s conclusion with respect to this study and similar ones is: “The fact that participants nevertheless came to view the person whom they wanted to like and respect as especially likable and competent in these studies therefore provides strong evidence for the role of motivation in judgment” (Kunda 1999, p. 216).

Other studies provide additional evidence for motivated reasoning (see Kunda 1990). It has been found that if a certain person finds an event desirable, he will believe that the event is much more likely to occur than if he had found the event undesirable. It has also been found that people tend to take credit for their success but to deny responsibility for failure.

<sup>11</sup> This study is discussed in Kunda (1990, p. 486; 1999, pp. 215-216).

Suppose now that we accept Kunda's conclusion,<sup>12</sup> so suppose there is the phenomenon of motivational reasoning, i.e. that the motivation to arrive at a particular conclusion affects the processes operative in belief formation. Where does that leave the thesis that belief is involuntary and the claim that "paying selective attention to the evidence" and the like are not policies one can adopt in order to secure that one winds up believing the specific proposition *p*?

The first thing to say is that the phenomenon of motivated reasoning does *not* show that belief is under immediate voluntary control after all. It doesn't show that we can decide, by *fiat*, to believe just any proposition *p* we like. Nor does, secondly, the phenomenon of motivated reasoning show that "paying selective attention to the evidence", contrary to what I have been arguing, *can* be adopted as a policy to ensure a wanted belief – it *doesn't* show that belief *is*, after all, under long-range voluntary control. The crucial point here is that motivated reasoning, just as "paying selective attention to the evidence", is a highly unconscious phenomenon. (We may, of course, observe that others engage in motivated reasoning, and we may note that others pay selective attention to evidence, but we usually don't observe that in ourselves.) And exactly this feature prevents them from being policies that can be adopted so as to secure long-range voluntary control over one's belief that *p* (if it is *p* one wants to believe). Condition (B) for voluntary belief fails to be met by them.

Thus far I have broached empirical psychological studies that, at a first glance at least, might be taken to counter my claim that belief is voluntary. I have argued that these studies do no such thing. I now turn to psychological studies that might seem to *support* the thesis that belief is involuntary.

Possible support comes from research in which the concept of "the unconscious" plays a crucial role. That research indicates that we are unconscious of much of what we see and feel, but also, more pertinent to the topic of this paper, that we are unaware of at least many of our beliefs and of the processes that propel them. My claim is that this research supports my claim that belief is involuntary. The argument for it is this:

- (1) What one is inescapably unconscious of is not under one's voluntary control.
- (2) Belief as well as belief formation is often inescapably unconscious.
- (3) Therefore, unconscious belief and unconscious belief formation are not under one's voluntary control.

<sup>12</sup> An important part of Kunda's work is to delineate the mechanisms that underlie motivated reasoning. One such mechanism is "biased memory search".

It is clear that this argument, if successful, at best *partially* supports my claim. After all, my claim is that belief is involuntary (by which I mean that we neither have immediate nor long-range voluntary control over a specific belief), whereas the conclusion of the present argument is only that *unconscious* belief formation is involuntary, and thus doesn't address the possibility that conscious belief formation is voluntary. Furthermore, so far I have been assuming that believing is a conscious affair: if one believes that *p*, then one is conscious of the fact that one believes *p*. Still, the argument, if successful, supports the conclusion of the previous section. For it might very well be that processes that lead up to belief are themselves oftentimes unconscious. My discussion of this argument consists of a discussion of the premises, from which the conclusion self-evidently follows.

The first premise states that what one is inescapably unconscious of, is not under one's voluntary control. What does it mean for one to be "inescapably unconscious" of something *X*? It means that it is impossible for one to be conscious of *X*. There are many things that one might be unconscious of, but that one is not *inescapably* unconscious of. You drive your car and the car is under your control. You are not consciously aware of the many intentional movements you make with your feet and hands, nor of all the individual cars on the road. But you are not *inescapably* unconscious of all of this, for it is not impossible for you to be consciously aware of these items. Had the first premise been "What one is unconscious of, is not under one's voluntary control", it would have been obviously false. For driving your car, as indicated, is in many respects an automated and unconscious affair; nonetheless, driving your car is under your voluntary control. But if you were to be *inescapably* unconscious of all the items I have mentioned, driving your car would *not* be under your voluntary control. I am inescapably unconscious of the growth of my hair and nails; that is why I have no voluntary control over the growth of my hair and nails. That is the substance of the first premise. If, as I shall assume, the first premise is true, this means that there is no point in talking about "inescapably unconscious control". Control over *X* can be unconscious, but it cannot be *inescapably* unconscious. Applied to belief this means that if one is to have control over the formation of belief, it is required that one is not *inescapably* unconscious of the processes that propel it.<sup>13</sup>

The previous paragraph was mainly a comment on the "inescapable" part of "inescapably unconscious". The next paragraph, which starts the defence

<sup>13</sup> This intuition is shared by Kunda; see Kunda (1999, p. 268).

of premise 2, is mainly a comment on the “unconscious” part of that expression.

The second premise is a summary statement of some findings of empirical psychology. It is, furthermore, an empirical, as opposed to a conceptual, statement. Now what does “unconscious” in this premise mean? What is it for something to be “unconscious” in the sense intended? Wilson says: “A simple definition of the unconscious is anything that is in your mind that you are not consciously aware of at a particular point in time” (Wilson 2002, p. 22).<sup>14</sup> He avows, however, that this definition is too simple because it “stretches things”. It would imply, for instance, that someone who isn’t currently thinking about his hometown is unconscious of “Philadelphia”, if that is his hometown – which doesn’t seem right. The problem with the simple definition (which has this unwelcome implication), he says, is that it equates consciousness with attention or short-term memory, while there are things one isn’t currently paying attention to, or aren’t currently in one’s short-term memory that still aren’t “unconscious” in the preferred sense. He therefore says “A better working definition of the unconscious is *mental processes that are inaccessible to consciousness but that influence judgments, feelings, or behaviour*” (*ibid.*, 23), but acknowledges that the unconscious is notoriously hard to define and that this is only one definition out of many.<sup>15</sup> He even says that he doesn’t want to get bogged down by definitional issues and that it is more interesting to take a look at what humans can accomplish outside the spotlight of consciousness. Still, it merits taking a closer look at the definition and bringing out certain problems. (i) The definition restricts the unconscious to “mental processes”, but, as the literature on the unconscious makes abundantly clear, not everything that “is part of” one’s unconscious is a *process*; there are, it is claimed, unconscious feelings, and, more relevant for my topic, unconscious beliefs – but neither feelings nor beliefs are processes. (ii) The definition says that what is unconscious is “inaccessible” to consciousness. But this can’t be correct, for the research of the unconscious is intent on studying it and hence on bringing to consciousness how the unconscious works and what it does (and even how one can get to know one’s unconscious). Perhaps the definition needs to be taken as affirming that subjects can have no access to their *own* unconscious – although they can have it to the unconscious of others (especially when they are empirical psychologists). But this cannot be right either, for in the same

<sup>14</sup> Wilson (2002, ch. 1) makes it clear that the modern notion of “the unconscious” is very different from Freud’s in that it has nothing to do with repressed infantile thoughts that are kept out of consciousness because they are a source of psychic pain.

<sup>15</sup> Miller (1942) offers no less than sixteen different definitions of the unconscious!

book in which he offers the definition, Wilson aims to provide clues as to how one can get to know at least parts of one's own unconscious.<sup>16</sup>

So, if the working definition doesn't work, there is no alternative to giving examples of the unconscious at work. The argument is limited to unconscious belief and the formation thereof<sup>17</sup> and so I will give examples of that. One example is due to the French physician Edouard Claparede and concerns a woman suffering from amnesia. Every time she sees Claparede, she wouldn't have recollections of previous consults and accordingly he would have to introduce himself anew each visit. One day Claparede decided to shake hands with the patient (as he always did) but this time while holding a hidden pin in his hand. The woman felt, of course, a painful prick and withdrew her hand immediately. What is interesting is what happened during the next consult. Again the patient showed no sign of recognizing Claparede, but when Claparede reached out his hand the patient refused to shake it!<sup>18</sup> One way to interpret this behaviour is to say that although the patient did not consciously recognize Claparede, and accordingly didn't have any conscious beliefs about him, she unconsciously believed that shaking hands with this man (who is in fact Claparede) might give her a painful sensation.

What I have described in terms of an unconscious belief that the patient acquired, can also be (and is in fact by psychologists) described in terms of unconscious learning. These two descriptions, however, aren't at odds with each other in that believing and learning seem to be closely connected with each other. When someone learns that *p*, he will usually thereby come to believe that *p* – whether the learning is conscious, or unconscious.

Another example of unconscious learning is described by Pawel Lewicki c.s.<sup>19</sup> In a laboratory situation participants were asked to watch a computer screen that was divided in four quadrants. On each trial, the letter *X* appeared in a quadrant, and the participants' task was to press one of four buttons to indicate in which of the quadrants the letter *X* appeared. Unbeknownst to the participants, the presentation of the *X*'s followed a complex rule. As time went by, their performance improved, they became faster and faster at pressing the correct button, and hence they appeared to learn the rule. But they were unable to verbalize, or even notice, it. After a certain

<sup>16</sup> See especially the later chapters of Wilson (2002).

<sup>17</sup> The unconscious is also said to be at work in, among others: proprioception; the way light rays strike the retina and are transformed into three-dimensional vision; filtering of information; interpretation of behaviour; evaluation; goal-setting.

<sup>18</sup> This case is described by Wilson (2002, p. 25).

<sup>19</sup> Lewicki, Hill and Bizot (1988); this is discussed by Wilson (2002, pp. 26-27).

elapse of time (and still unbeknownst to the participants), the researchers changed the rule with the result that the participants' performance deteriorated dramatically. Although the participants noticed this, none of them knew why. And this shows, or so argue Lewicki c.s., that the participants had learned the original rule *unconsciously*, and hence also knew (and believed) it unconsciously.

A further, and widely recognized, example of unconscious learning and belief is language learning, which, in young children, goes almost without effort and without conscious awareness of what has been learned. Small children learn the syntax, grammar and semantics of a language without consciously learning syntactical, grammatical and semantic *rules*. They are typically unable to even explicate them. Still, they would be able to identify syntactical, grammatical and semantic mistakes – which suggests that they know (and hence believe) the relevant rules, albeit it unconsciously.

So far, the examples were of unconscious belief – beliefs of which people are supposed to be unconscious. In addition to this, research also indicates that many processes that trigger conscious belief escape the light beam of consciousness. One famous experiment that suggests this is due to Robert Nisbett and Timothy Wilson. Passengers in a shopping mall were asked to examine the quality of four pairs of nylon hoses that were neatly arranged on a table and say which pair was of the best quality. As a matter of fact the pairs were of equal quality. But 40% of the respondents said that the rightmost pair was the best, and 31% the pair second from right. Upon being asked why they believed that the pair they selected was best, the respondents would always point to an attribute of the preferred pair, such as its superior knit, sheerness, or elasticity. What is at work here, Nisbett and Wilson argue, is a so-called “position effect” – in the case the effect that people have a marked preference for items on the right side of a display. None of the respondents, of course, was conscious of this effect, and they rejected the very suggestion that their belief might be due to such an effect. Nonetheless, the effect is real, or so these studies argue.

Other research indicates that belief formation is often automatic in the sense that we are not aware of the cognitive processes that underlie it. Numerous studies indicate the unconscious effects of priming. A summary statement of such research by a renowned psychologist reads:

Much of our mental life goes on automatically. Unlike controlled processes, automatic processes occur outside awareness, are carried out without intention, are uncontrollable (i.e. one cannot stop them), or are highly efficient (i.e. require few cognitive resources). Our judgments [beliefs, RvW] . . . can be influenced by factors that we have never been aware of and have only been exposed to subliminally . . . and by factors that we can still recall but whose influence we are unaware of. (Kunda 1999, p. 308)

Hence the conclusion of the argument stands: inescapably unconscious belief and inescapably unconscious belief formation are not under one's voluntary control. And this, as I have explained earlier on, in turn supports, at least in part, the claim that belief is involuntary – the claim that I set out to defend in the previous section.\*

Department of Philosophy  
VU University Amsterdam  
The Netherlands

E-mail: r.van.woudenberg@vu.nl

### References

- Alston, W.P. 1989: "The Deontological Conception of Epistemic Justification", in Id., *Epistemic Justification. Essays in the Theory of Knowledge*, Ithaca, Cornell University Press, pp. 115-152.
- Alston, W.P. 2005: *Beyond "Justification". Dimensions of Epistemic Evaluation*. Ithaca, Cornell University Press.
- Berscheid, E., Graziano, W., Monson, T. and Dermer, M. 1976: "Outcome Dependency: Attention, Attribution, and Attraction", *Journal of Personality and Social Psychology*, 34, pp. 978-989.
- Butler, J. 1889: "Upon Self-deceit", in Id., *The Analogy of Religion, Two Brief Dissertations, and Fifteen Sermons*, London, George Bell and sons, pp. 473-484.
- Edgley, R. 1969: *Reason in Theory and Practice*, London, Hutchinson.
- Feldman, R. 2001: "Voluntary Belief and Epistemic Evaluation", in M. Steup (ed.), *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, Oxford, Oxford University Press.
- Kunda, Z. 1990: "The Case for Motivated Reasoning", *Psychological Bulletin*, 108, pp. 480-498.
- Kunda, Z. 1999: *Social Cognition. Making Sense of People*. Cambridge, Mass.-London, MIT Press.
- Lewicki, P., Hill, T. and Bizot, E. 1988: "Acquisition of Procedural Knowledge about a Pattern of Stimuli that Cannot Be Articulated", *Cognitive Psychology*, 20, pp. 24-37.
- Miller, J.G. 1942: *Unconsciousness*, London, Chapman & Hall.

\* For comments on earlier drafts of this paper my thanks go to audiences in Paris, Louvain, Edinburgh, and Bologna, to Robert Audi, Stefaan Cuijpers, Lieven Decock, Igor Douven, Anders Kraal, Rik Peels, Duncan Pritchard, Jeroen de Ridder, Emanuel Rutten, and William Wood. It is thankfully acknowledged that work on this paper was supported by a Plantinga Fellowship at the University of Notre Dame.

- Peels, R. 2012: *Believing Responsibly. Intellectual Obligations and Doxastic Excuses*, Utrecht, Quaestiones Infnitae. Publications of the Department of Philosophy. (Dissertation)
- Shilts, R. 1987: *And the Band Played on: Politics, People, and the AIDS Epidemic*, New York, St. Martin's Press.
- Weatherhead, L.D. 1951: *Psychology, Religion, and Healing*. New York-Nashville, Abingdon.
- Wilson, T.D. 2002: *Strangers to Ourselves. Discovering the Adaptive Unconscious*, Cambridge, Mass.-London, Belknap Press.
- Winters, B. 1979: "Believing at Will", *Journal of Philosophy*, 76, pp. 243-256.
- Wolterstorff, N. 2010: *Practices of Belief*, Cambridge, Cambridge University Press.
- Woudenberg, R. van 2009a: "Responsible Belief and Our Social Institutions", *Philosophy*, 84, pp. 47-74.
- Woudenberg, R. van 2009b: "Ignorance and Force. Two Excusing Conditions for False Beliefs", *American Philosophical Quarterly*, 46, pp. 373-386.