The Ethics of Alien Attitudes

1. Introduction

People do not always really believe what they take themselves to believe. A person may sincerely say that a certain racist belief is definitively false, but still hold such a belief. When asked what he believes about something, it is likely that he simply expresses his opinion about the issue in question, and this reveals what he takes himself to believe, but not necessarily what he *really* believes. In some cases, however, a person may adopt a kind of a third-person point of view. Instead of expressing his opinion, he may *report* the attitude he has in light of convincing evidence concerning his behavior and other attitudes. It follows that sometimes a person may report having an attitude which conflicts with his better judgment (his opinion). "Many people have completely unjustified racially biased beliefs and, judging from my behavior I must admit that I have them myself." In these cases the person's (evidential) beliefs are not apparent to the person in the normal way, and are not judgment-sensitive (or reason-responsive) in a way that they are supposed to be.[1] Beliefs of this kind can be called *alien attitudes,* or more narrowly, alien beliefs. They are attitudes or beliefs that fail to be sensitive to the person's regular processes of introspection and evaluation and are known by him merely through behavioral and psychological evidence that he has noticed about himself, or learned about himself from others.[2] When a person is aware of his beliefs in this way, he is not committed to their truth or overall acceptability; he has not endorsed them as true.[3] To *have* an alien attitude of this sort is to *realize* that one has conflicting attitudes, with some (the "alien" ones) being very oddly related to oneself. Most or all people have unnoticed beliefs that conflict with their sincere opinions, but the unnoticed beliefs of this sort are not "alien" in the relevant sense, as they *do not appear as alien* to those who have them.

"The Ethics of Alien Attitudes" by Juha Räikkä and Saul Smilansky,
The Monist, vol. 95, no. 3, pp. 511–532.

In daily life people are typically interested in others' opinions. When people say "I love you," for example, or make contracts, they express their opinions, they do not merely report their beliefs by referring to the best evidence available, and this is also what they are expected to do. There are some exceptions to this. Employers, for instance, may make use of psychological tests in order to find out jobseekers' evidential beliefs (and character traits) instead of their stated opinions, however sincere they may be. However, in general, opinions suffice in everyday life—maybe because people assume (wrongly) that others' explicit opinions accurately reflect their evidential beliefs, or because they are not interested in others' evidential beliefs in the first place. It is often much more important to know how people justify their choices than to learn what really influenced them when they made those choices and what beliefs should be attributed to them.[4] If we wish to find out whether a person speaks sincerely, say, in a court of law, we need to determine whether what he says is what he *takes* himself to believe, *not* whether what he says is what he really believes, in the evidential sense of believing.[5] Of course, it may be important to know what a person's evidential beliefs are, because they tend to influence his behavior. But evidential beliefs *need not* influence behavior. When a person who sincerely says that racism is wrong hears from a reliable source that in fact he has racially biased beliefs and accepts the disappointing news, he can try to make sure that those biased beliefs do *not* affect his behavior. To a certain degree, it depends on the person how he deals with his alien attitudes.

In what follows, we aim to analyze some ethical dilemmas that people may face when they deal with their alien attitudes, in particular, their alien beliefs. We will ask what might follow if alien attitudes are treated as a part of "reality," and what kinds of ethical problems can arise when a person wishes to identify with his or her alien attitudes. To an extent, the dilemmas that alien attitudes provoke are familiar from other segments of our psychological lives. This is not surprising as alien attitudes are not the only judgment-sensitive attitudes that fail to be sensitive to judgments in the required way. For instance, when a person feels "irrational guilt," his judgment-sensitive attitude, guilt, appears to be inconsistent with his judgment that he has done nothing wrong (hence irrationality).[6] Typical irrational guilt, however, is not an alien attitude because it is observed (i.e., felt) by the person in an immediate first-person

way—not through assessment of evidence about himself. Up to a point, alien attitudes raise ethical problems of their own, or so we will argue.

This paper is motivated by the assumption that the advanced neurosciences that have revolutionized the empirical study of the human mind may lead to a situation where having alien attitudes—being aware of one's conflicting beliefs—will be a much more common phenomenon than it is at present. There will be more and more applications in the fields of functional MRI (Magnetic Resonance Imaging) or other such methods, and brain imaging will provide people with various kinds of information about their inner lives, including information that they have many evidential beliefs that conflict with their seemingly sincere opinions. To some extent, this has already happened.[7] Although the ethical dilemmas that alien attitudes generate may not be characteristic problems of "neuroethics" (related to wrong diagnoses, false hopes, free will, etc.), it is fair to say that if people will need to confront the realization of a growing number of alien attitudes in the future, this will happen mainly because of the groundbreaking research done by the neuroscientists. It is often said that a person's true secrets are more secret to him or herself than they are to others, but it is unclear how long we will have such secrets.[8]

2. Reprehensible, Neutral, and Morally Desirable Alien Attitudes

Let us start by distinguishing reprehensible, neutral, and morally desirable alien attitudes.[9] From an ethical point of view, the most interesting categories are reprehensible and morally desirable. The issue of alien attitudes is a part of the larger issue of conflicting attitudes.

Prejudices and biases that people tend to carry with them form the basis for reprehensible alien attitudes. Racist and sexist beliefs are typical examples of such attitudes, but these need not of course target another race or sex. The target can be another country, or residents of a particular community. Sometimes there is no special "target" at all. Reprehensible alien attitudes are alien to a person in the sense that they are not his opinions and are known by him merely through evidence that he has noticed himself or learned from others. Notice that a person who often entertains racist thoughts and who blames himself because of them is *not* dealing with alien attitudes. He has immediate cognitive access to his racist thoughts, and his awareness of them is not based on any sort of combination of reliable evidence and deduction. For our purposes, reprehensible alien attitudes can be taken to

be reprehensible in the sense that both he and a number of other people around him think that the attitudes are morally bad and that he should not harbour such attitudes. The attitudes are condemned on the grounds that they tend to generate wrongful action or because they indicate an undeveloped moral character and are condemnable as such, whether or not they have any further consequences. (For the sake of this discussion, we shall bracket the question whether reprehensible attitudes are reprehensible *in fact*, and assume that this is not in doubt.)

Neutral alien attitudes are familiar from experimental psychology. For instance, in their famous review article "Telling More Than We Can Know: Verbal Reports on Mental Processes" (1977) Richard E. Nisbett and Timothy D. Wilson report on a study in which they asked fifty-two subjects to evaluate four *identical* pairs of nylon stockings that were arranged, in a line, in front of them. Subjects were asked to say which stockings were of the best quality and why. The stocking to the right was heavily overchosen: the right-most stockings were preferred over those to the very left by a factor of almost four to one. However, none of the subjects justified the choice by referring to the position of the article in the line, and even when they were asked to consider the possibility that the position was the criterion they used, "virtually all" of them denied that it was.[10] This study is clearly relevant to alien attitudes. Suppose that after the study the participants concluded that, strangely enough, they seemed to have the belief that "The right-most stockings are better than the left-most stockings because of their location" (for otherwise their choice is unexplicable).[11] In this case they now realize that they have had very odd beliefs, which now become alien attitudes.[12] However, alien attitudes of this type can be called "neutral," as it is unlikely that anyone would think that having them is particularly bad or good, from a moral point view. Research in experimental psychology suggests that people may have excellent justifications when choosing how to think or act, but still make their choices on the basis of reasons other than those they would refer to if asked. The same general result is the main message of the recent work on "adaptive" unconsciousness: "we (more often than we might have thought) perform poorly in judging the causes and reasons of our actions and behaviours."[13]

A person may have a bad conscience or guilt feelings just because he does not believe something he thinks he, as a moral person, should believe. These feelings open the door to what we will call morally desirable alien attitudes. Consider an example. A priest may fail to judge that God

exists, although he feels that he should believe in God. He sees his inability as a moral flaw, and so does the community around him.[14] Suppose, however, that a group of talented theologians prove that actually he does believe in God, in the evidential sense of believing, and tell the good news to our priest, who then becomes convinced that actually he does believe. Now he has an alien attitude which is morally desirable. Morally desirable alien attitudes are alien for a person in the same sense as reprehensible alien attitudes can be alien for a person. The beliefs in question are not a person's "conscious" opinions and are known to him merely through reliable evidence that he has noticed himself or learned from others. Morally desirable alien attitudes are desirable in the sense that both the person himself and relevant people around him agree that the attitudes are morally good and that, if possible, he should hold such attitudes. Notice that such attitudes need not be (and usually are not) connected to religious beliefs. An environmental activist may fail to judge that "My daily choices concerning how I move from one place to another are important in the fight against global climate change." But if he in fact believes this, then he has a morally desirable alien attitude, i.e., if he is aware of his evidential belief and thinks (correctly perhaps) that having it is morally important.[15] Notice also that the attribution of evidential beliefs can be particularly complicated in the case of morally desirable alien attitudes. A priest who cannot accept that God exists may still intentionally behave *as if* he believes that God exists, and his behaviour can cause mistaken belief-attributions. (This is why the group of theologians who make the belief-attribution need to be talented.) It may seem surprising that people may have alien beliefs that are better than their conscious beliefs, since it is easier to understand how hiding reprehensible beliefs can be more natural and functional than admirable ones. But as our examples show, alien attitudes, especially beliefs, can be favourable.[16]

How people should deal with their alien attitudes is an important question. Morally reprehensible and morally desirable alien attitudes both raise ethical dilemmas. Let us start by considering alien attitudes that are considered moral deficiencies.

3. Alien Attitudes as a Part of "Reality"

One way to deal with alien attitudes is to treat them in an "I have a broken arm" manner. A person who has a broken arm needs to take this fact into account in his decision making. Having a broken arm may have

considerable effects on his decisions concerning traveling, shopping, and so on. Obviously, a person *can* look at his alien attitudes in the same fashion as he looks at his broken arm. Whether it is ethically unproblematic to do so is another matter. Let us consider a concrete example. In *Being Known* (1999) Christopher Peacocke describes a case of a person (we assume she is a professor) who is biased against people who have undergraduate degrees from countries other than her own:

> Someone can make a judgement, and for good reasons, but it not have the effects that judgements normally do—in particular, it may not result in a stored belief which has the proper influence on other judgements and on action. A combination of prejudice and self-deception, amongst many other possibilities, can produce this state of affairs. Someone may judge that undergraduate degrees from countries other than her own are of an equal standard to her own, and excellent reasons may be operative in her assertions to that effect. All the same, it may be quite clear, in decisions she makes on hiring, or in making recommendations, that she does not really have this belief at all. In making a self-ascription of a belief on the basis of a conscious judgement, one is relying on the holding of the normal relations between judgement and belief which are not guaranteed to hold.[17]

Suppose now that the professor's colleagues convince her that in light of well-established and intuitively plausible principles of belief attribution, she does not really believe that undergraduate degrees from countries other than her own are of a standard equal to her own, given her behavior in hiring and making recommendations. Instead, her behavior suggests that she believes that it is not the case that undergraduate degrees from countries other than her own are of a standard equal to her own.[18] The result is an attitude that clearly conflicts with her (at least apparently) sincere judgment that undergraduate degrees from foreign countries *are* of a standard equal to her own. Her biased evidential belief—"undergraduate degrees from countries other than my own are not of a standard equal to my own"—is completely alien to her. She is aware of the belief merely through external evidence.[19] Although she may be proud of her national sentiment in other contexts, she finds it uncomfortable to admit that her nationalism manifests itself in such a deplorable form.

Let us assume that the principles of belief attribution used were correct and that the professor really has an alien attitude, a belief.[20] Let us also assume that the alien belief is plainly false (as she also thinks), and she realizes that this has caused her to make unfair and discriminatory

decisions. What should she do? Presumably, her colleagues would expect her to get rid of her bias, now that she is aware of it. Indeed, some of the colleagues may think that while she was not previously responsible for holding such an attitude, now she is, because she is aware of its existence and of this fault in her character.[21] If the professor is fortunate, she will get rid of her biased belief—perhaps simply by asking herself the deliberative question "What am I to believe?"[22] Because our beliefs are not always formed as a result of explicit deliberation, engaging in deliberation may change beliefs, including evidential beliefs. However, things can be much more complicated. Merely repeating something is not a particularly effective way to free oneself from prejudices that one finds ridiculous.[23] Consider morbid jealousy. A person who suffers from morbid jealousy wants to free himself from the feeling, and a therapist can help him to see that it is clearly unfounded. But it may well happen that his painful emotional state does not change, no matter how hard he tries. The same is true about the alien attitudes. One's disapproval of those attitudes does not always imply that they disappear. The connection between the effort and the improvement is insecure and random.

Compare the professor's situation to that of a man who has a stubborn "intuition" that "undergraduate degrees from foreign countries are not of a standard equal to my own" but considers that intuition false and wants to get rid of it. The man has immediate (as opposed to merely evidential) access to his biased attitude—as does a person who suffers from morbid jealousy has immediate access to his painful feelings. The man with the intuition is in a better position than the professor, as he can at least analyze his attitude with his inner eye, describe it to his friends who may want to help him, tell when it occurs, characterize its strength, and so on. The professor seems to be in the dark with respect to her alien attitude. Both of them may fail in their attempts, but the prospects of the man with a "strong intuition" are more promising, perhaps much more promising. He is in touch with his (biased) emotions and beliefs in a direct way, and has access to them when he puts his mind to it.

If the professor fails to get rid of her biased alien belief, she may try to prevent it from influencing her decisions by taking precautionary measures. She may well feel that she has a duty to do so. This strategy is an obvious alternative, although it may frustrate some of her colleagues. They may feel respect for her, for trying to free herself from her prejudice, yet the

result is a disappointment—whether or not she is blameworthy because of the result. It is a disappointment because one feels that a moral person should be able to overcome such biased attitudes, and not only merely arrange matters so that, while her biased attitudes remain, they are no longer effective. The second strategy suggests that, in a sense, she has given up on her alien attitude. At least, she seems to confess that the first strategy may not be successful in the near future, and treats her alien attitude as a part of "reality," although she understands perfectly well that the attitude is faulty, since the content of the belief is false. Her confession that she may be unable to get rid of her bias may make it less likely that she *will* get rid of it. Now she probably works less to get rid of it. Therefore, the launch of the second strategy is not ethically unproblematic.

The precautionary measures she takes may include a decision to be especially careful when she deals with people who have undergraduate degrees from foreign countries, or a decision to consult her colleagues in such cases. *If* she succeeds in stopping her discriminatory behavior, this does not mean that she has got rid of her biased belief.[24] The alien attitude can still be attributed to her, for instance, on a counterfactual basis: had she not taken the precautionary measures, she would have discriminated against people because they had undergraduate degrees from countries other than her own. However, experience yields no guarantee of success for this second strategy, either. Precautionary measures may work well, but they can also work only partly or fail completely. Prejudiced people who fight against the overt consequences of those prejudices can be quite unsuccessful, and sometimes their efforts lead to incidents that have elements of farce. If the professor starts to think that she cannot free herself from the biased belief and that she can prevent its influences over her behavior only partly—not an unlikely scenario—she certainly feels that her position is unsatisfactory both epistemically (her alien belief is false) and morally. In fact it is likely that whatever precautionary measures she may take, over time the alien beliefs will manifest themselves in some way, bypassing the external defences set against them.

However, the professor can apply a third kind of strategy to deal with her alien attitude. Suppose that she reasons as follows: "It is likely that I cannot get rid of my stupid attitude. It is part of my mental furniture. Because I am biased against people who have undergraduate degrees from foreign countries, it is likely that I will be a better colleague and teacher

if I work with people who do not have undergraduate degrees from those countries. Therefore, it is better if I continue to make my recommendations and decisions on hiring in the way I am used to. After all, it is likely that I will have to work with the people whom I recommend or hire." This is an argument in favor of discrimination, but it is not an argument for the view that it is acceptable to defend one's decisions with a discriminatory criterion, such as the national origin of the undergraduate degrees. She has not defended her decisions by such a criterion. Rather, it is a pragmatic strategy: given that her attitudes towards a certain class of persons are in fact unlikely to change, it is in some sense better if she does not work with such people.

When the professor tells about this new move to her colleagues, they are not likely to be happy about it. Her defense of her old practices on the grounds that they will have desirable consequences is highly dubious, even if it is correct that in consequentialist terms this may indeed be the best way to proceed. The colleagues think that it is much more important to avoid discrimination than the professor seems to assume, although avoiding it may have some moral costs in the professor's case.[25] The colleagues also point out that the professor resembles an addict who would prefer not to have a desire to smoke but justifies his smoking by referring to the relaxing effects of satisfying the unwanted desire. A difference between the professor and the addict is that the professor could act otherwise while the addict could not. The professor could apply her second strategy and try to prevent the alien attitude from influencing her behavior. Indeed, it is up to her whether she follows the first, second, or third strategy in dealing with her alien attitude. But all of them can be problematic, as the discussion above shows.

In "Involuntary Sins" (1985) Robert Adams argues that "the struggle against a wrong state of mind in oneself is normally a form of *repentance*, which involves self-reproach."[26] Adams writes that at the "center of such a process is one's *taking responsibility* for one's state of mind" and that when you "take responsibility for it you also do not see it as something that just happens to you, like a toothache or a leak in your roof."[27] Perhaps this is what "normally" happens (Adams' example is a person who has "just realized" that he is "ungrateful to someone" who has done a lot for him), but it is not what happens when a person notices merely on evidential grounds that he has an attitude that is clearly reprehensible. A person may

have a close relation to his toothache or his broken roof, at least if he thinks that it is partly his own fault that he has a toothache or a broken roof (as is often the case). But when a person struggles against his alien attitude and "takes responsibility" for it, this is probably not because he realizes that it is his own fault that he has it, but because he feels that it is still in some sense part of him, and that he is the only person who could perhaps eradicate it. One can feel responsible for dealing with some manifestation of oneself, even if one is not, in fact, at fault for it.[28] And this sense of responsibility, as opposed to what Adams seems to say, is *compatible* with the idea that what he is responsible for is, in his eyes, something that just happened. If a person does not manage to get rid of his reprehensible alien attitude and has no idea of its origin, he may incredulously wonder, "What is it doing there?"[29] If he thinks that he has it because of a bad upbringing or an unfavorable social environment, he may bitterly ask, "Why has this happened to *me*?"

4. Identifying with Alien Attitudes

So far we have discussed false and reprehensible alien attitudes. But alien attitudes can also be morally desirable, and in these cases people may *want to identify with them*. Of course, a person cannot identify with his alien attitude in the sense that he could directly use it as a premise in his reasoning about what to think or do.[30] When a person reflects on how to act or what to think, he must rely on premises he takes to be true. "Intuitions" that fail to be judgment-sensitive *can* be used as justificatory reasons because it is still possible to take them to be true—even when one is unable to provide a justification for them. If a person's overall reasons conflict with his strong intuition, he can always think that he must have missed something in his reasoning and keep on believing that the intuition is true.[31] But a person who has an alien attitude does *not* take his attitude to be true (as he takes his nonalien beliefs to be true) and therefore cannot identify with the alien attitude in this strong sense. However, it is possible to identify with one's desirable alien attitudes in other ways. For instance, a person (1) may want to think that he is "really" the kind of person that his alien attitude suggests, or he (2) can express his alien attitude when he is asked what he believes. This would not necessarily mean that he is a hypocrite or a liar, although a modicum of self-deception or at least wish

fulfillment would probably be involved. Let us examine both of these ways of identifying with one's alien attitudes and explore their ethical status.

Consider the case of the chief executive officer of a large company who fails to judge that "women are as able as men to act in responsible leadership positions in business enterprises." The CEO has had bad experiences with women leaders in the firms he has headed, and even when he considers the issue very carefully, repeatedly, he is unable to change his mind. This makes him unhappy. He is well aware that clever people in his firm and elsewhere tend to disagree with him on this matter, and the worst thing is that his failure to share a more egalitarian view is commonly seen as a moral flaw. The CEO agrees that it *is* a moral flaw. He is not a chauvinist and understands without any difficulty that, morally speaking, he *should* think that "women are as able as men to act in responsible leadership positions in business enterprises."[32] Having this sort of belief is required in order to express a civilized attitude and a good moral character, and may prevent sex-based (gender) discrimination in the workplaces. But he cannot just *decide* to form this belief—as he does not feel that he has sufficient evidence for it, and remembers all too well his disappointing experiences with female leaders. These experiences are simply too salient for him. Suppose, however, that a group of top psychologists who (for some reason) interview him (or scan his brain) prove that actually he *does* believe that "women are as able as men to act in responsible leadership positions in business enterprises," in the evidential sense of believing.[33] The group tells the great news to the CEO who feels enormously relieved—despite his "schizophrenic" situation.[34] He is now aware that he *has* the politically correct and morally desirable belief that "women are as able as men to act in responsible leadership positions in business enterprises" (although he still thinks that there are insufficient grounds to think that the view is true).

In these circumstances the CEO may want to identify with his alien attitude and think that he "really" is the kind of person that his alien attitude suggests. As the alien attitude is morally desirable and socially beneficial, it is only natural to expect that he would do so. He may reason as follows: "A person who has a *reprehensible* alien attitude may feel guilt because of it and can take responsibility for it, because the attitude is truly *his* attitude, in the sense that he feels that he is the only person who could,

even if only in principle, eradicate it.[35] But if reprehensible alien attitudes are proper sources of guilt and shame, then morally desirable alien attitudes can be proper sources of contentment and pride. I am really a person who opposes sex-based discrimination in work places. I am not like my chauvinist ancestors."

However, this kind of self-understanding can itself be the object of moral evaluation, and it is far from clear that the self-understanding of the CEO as described above is ethically without problems. It is psychologically unlikely that the CEO can simply accept the fact that he has a morally desirable alien attitude that conflicts with his opinion. What is likely to happen when he hears about his alien attitude is that he *starts to reassess* the issue whether women are as able as men to act in responsible leadership positions in business enterprises. He cannot just smile and confess that he has an evidential belief that is in obvious contradiction with his explicit opinion based on his past experiences.[36] Perhaps he could try not to think about the issue, but this would mean that he could not praise himself with the view that he is really a person who opposes sex-based discrimination in work places. Such a thought would open the question again, with the possible result that his morally desirable alien attitude suddenly disappears. That would not be a desirable consequence.

If the CEO is fortunate he will find sufficient evidence for the view that women can be as able in business as men. Finding such evidence is not at all difficult, but appreciating it is difficult for *him*, given his past experiences and their salience for him. *If* he becomes convinced, the clash between his evidential belief and opinion dies out, his new self-understanding is unproblematic, and he can identify (in the strong sense mentioned above) with the attitude that formerly *was* alien to him. In these circumstances he may be tempted to accept a special version of epistemic conservatism.[37] The CEO may infer that his maintaining the evidential belief that "women are as able as men to act in responsible leadership positions in business enterprises" is itself evidence for the truth of that claim. "If I have such a belief, I must have formulated it on the basis of some good evidence. I have no idea what the evidence was, but it must have been there, because otherwise I would not have had such a belief in the first place." This move is clever, perhaps, but there is an air of self-deception now, especially if the existence of his alien attitude can be explained in a better way than by referring to the alleged evidence that

he once had (for instance, by referring to the prevailing social pressure to have egalitarian beliefs in gender issues).[38] Self-deception is often ethically problematic, and there is also no guarantee that the CEO will manage to change his mind by means of epistemic conservatism.[39]

Let us now turn to another way of identifying with morally desirable alien attitudes. Suppose that the CEO is giving his annual talk to the workers of the firm he is leading. As always, he announces that "women are as able as men to act in responsible leadership positions in business enterprises." He has been insincere before—in previous years he has said it merely in order to protect the firm's public image—but now his situation is radically different. He has (albeit in alien form) the belief that "women are as able as men to act in responsible leadership positions in business enterprises" and he is *aware of having it.* Is he still insincere when he makes the claim? Our answer depends partly on how the notion of sincerity is understood in this connection. In "Problems of Sincerity" (2005) Richard Moran defends the claim that, in a sense, the demands of sincerity are weaker than the demands of accurate presentation of one's beliefs and other attitudes but, in another sense, the requirement of sincerity is more demanding than the accurate presentation of one's state of mind. He describes the following case:

> For again, if someone has the repressed belief, for example, that he is a coward, but takes himself to believe no such thing, he will have failed to speak sincerely if, for his own reasons he nonetheless says that he *is* a coward, even though by hypothesis what he asserts here expresses what he actually thinks about himself. . . . Just as it is possible to lie while inadvertently reporting the actual facts, it is possible to speak insincerely while asserting what it is in fact one's actual belief. Saying what I actually believe is not *sufficient* for sincerity, if the belief expressed is not what I *take* myself to believe. And saying what I actually believe is not necessary for sincerity either, since I still speak sincerely if I am somehow wrong about my actual belief but nonetheless assert what I take myself to believe.[40]

This understanding of sincerity sounds plausible but it does not unambiguously tell us whether our CEO is sincere or not when he says to his audience that "women are as able as men to act in responsible leadership positions in business enterprises." This is because it is not clear whether a person who is aware of his beliefs merely through reliable external evidence (that he has noticed himself or learned from others) "takes" himself to believe the issues in question. The CEO certainly "takes" himself to *have*

the belief that he reports to have. This suggests that he speaks sincerely. But he does not take himself to believe that the belief he reports as having is *true*. He thinks (wrongly) that the claim that "women are as able as men to act in responsible leadership positions in business enterprises" is quite possibly incorrect. This is surely an odd and undesirable state of mind to be in, but it is not incoherent.

The assertion of the CEO is ethically problematic. Although he now takes the (we are assuming true) belief that "women are as able as men to act in responsible leadership positions in business enterprises" to be *his* belief and speaks in this sense sincerely, it is probable that his audience (or at least most of them) would like to hear the CEO expressing his *opinion* about the gender issue rather than only his *report* about his state of mind, however accurate the report may be.[41] *If* it is clear (or at least should be clear) to the CEO that the audience is not interested merely in such a report but rather in his opinion, a position that he is ready to defend, then, at first sight at least, he is an appropriate target for moral blame.[42] He is responsible for intentionally misunderstanding the rightful expectations of his audience, and seems to go along, misleading them about what is in fact the case. Here he seems open to the charge of insincerity.

Of course, in certain cases it is completely a person's own business whether he expresses his opinion or merely reports his beliefs and preferences. Voting is an example of such a situation. A person who has an alien attitude that "a conservative candidate is better than a liberal candidate," but whose reasoned opinion is that "a liberal candidate is better than a conservative candidate," is free to vote for whichever candidate he wishes. The members of a voter's "audience" may have some moral views on whether he should express his opinion or report his preferences—perhaps they would expect him to express his opinion[43]—but at the same time they are hardly willing to deny the voter's moral right to *choose* whether he expresses his *true* opinion or reports his *real* preferences. The CEO is not in a similar situation. His audience would like to hear him express his reasoned opinion about the gender issue, and the members of the audience do not think that it is none of their business whether he expresses his true opinion or merely reports what beliefs he seems to have, according to indisputable external evidence.

Another complication is that if the CEO has chauvinist (nonalien) beliefs there is perhaps a level at which it is good that he would express

in public only his alien beliefs, for politically correct reasons. But if we limit ourselves to the expectation for sincerity, then our CEO is not "saved" by reporting his alien beliefs as though they were simply "his beliefs." Perhaps he would do best by not speaking on the issue or, alternatively, by laying out all the cards, namely, sharing the predicament he finds himself in, with dubious beliefs but better alien beliefs vying for his endorsement in different ways. But while doing so may help him not to be insincere, it might well not be viable in practice if he wants to remain as a CEO of a public company. Having favorable alien beliefs does not really make his situation better, unless combined with insincerity about their alien nature.

As already suggested, sometimes people *are* interested in reports rather than opinions.[44] Whether a person should report his evidential beliefs or express his opinion is not always clear. These cases easily cause confusion, and the ethical acceptability of the speaker's choice depends, among other factors, on the expectations of his audience, on how justified those expectations are, and on the extent to which the speaker is and should be aware of those expectations.

Our focus is on the predicament of the moral person who finds herself laden with alien beliefs, which she then has to try to deal with, in particular in her representation of herself to others. There are metaphysical questions here ("Which is the real me?") and ethical ones, on which we have concentrated. In parallel to the self-related questions which we are exploring, there are also questions for others, such as how one is to judge people with (revealed) alien attitudes. But we are concerned here primarily with the first-person perspective. The nature of the alien beliefs and whether they are morally reprehensible or praiseworthy will play a major part in determining what one should do. But in addition to the concern with harming others (e.g., by choosing candidates while one is biased) and with sincerity (e.g., expressing one's only-alien beliefs about women business leaders as though they were "one's opinions") there is a whole range of further, broadly evaluative and normative concerns, focusing more directly on the agent herself. These are sometimes spoken of in terms of integrity (in the sense of wholeness), and sometimes in terms of authenticity. While recognizing the importance of these further issues, we have had to limit ourselves and leave these matters for another time. However, and just by way of gesturing to the importance of our

topic, it is clear that questions about integrity and authenticity will greatly increase in importance, to the extent that, due to technological developments, people will become much more aware of their conflicting beliefs.

The potential impact of an increased awareness of one's contradictory beliefs may perhaps be likened to the revolution brought forth by Freud's unmasking of the unconscious (and commonly infantile) basis for our outward reactions and behavior. If indeed we human beings consistently have alien beliefs alongside our explicit, conscious ones, and if new technologies will enable us to become much more aware of such divisions of the self, this should have significant repercussions for our self-image, and for the way in which we deal with ourselves and others. In particular, since it seems very likely that many reprehensible beliefs will be uncovered, and in an empirical way that will make deniability impossible, then we may confront a very difficult reality. People today are able to hide much of their mental life both from themselves and from others. A world with such enhanced self-awareness, emerging in "alien" form, and with much less privacy concerning one's beliefs appears extremely threatening.

Admittedly, just as many people react with interest to psychoanalysis, and feel that their lives have been deepened thanks to it, some people may be fascinated by the uncovering of their alien beliefs. Yet there are reasons to fear that here the situation may be less sanguine.

When one makes a "Freudian slip," the resulting understanding of oneself might be just as "alien" as in the sort of process with which we are concerned. However, the technological "unmasking" which we are exploring would be more systematic. By contrast, the more systematic Freudian processes, i.e., in psychoanalysis, may be just as disturbing, but they would not, typically, be as alien as those involving the fMRI (or similar technologies in the future). For, after all, in psychoanalysis the patient herself is doing most of the work, reporting on her dreams or associations, and gradually becoming aware of new aspects of her unconscious. There is something particularly stark about the idea of being confronted, quite suddenly, by an external empirical-scientific report that one holds such and such beliefs, beliefs that one had not realized. Moreover, the nature of the new technologies makes them more susceptible to widespread public use while psychoanalysis by its very nature is an intimate and private way of self-understanding.

5. Concluding Remarks

We have argued that alien attitudes may bring forth surprising ethical dilemmas for people who have them. Alien attitudes can be treated as a part of "reality," and that may cause problems. People can try to identify with their alien attitudes but this is also likely to lead to difficulties. Morally desirable alien attitudes do not seem less troublesome than reprehensible alien attitudes. Both of them are, at least potentially, problematic. In the discussion above we considered two examples (the case of the professor and that of the CEO), but it is unlikely that the ethical dilemmas would be very different even if the details of the cases changed. Whatever the concrete example, people who face reprehensible alien attitudes would normally like to get rid of them, and those who come to know that they have morally desirable alien attitudes face a natural temptation to make use of them. But even in the latter case the inner discord which is uncovered is disturbing.

As we mentioned at the beginning of the paper, this study is motivated by the assumption that the advanced neurosciences may make the realization that one has conflicting attitudes much more prevalent in the future than at present. Conflicting attitudes seem to be very common, and given the right technology, our awareness of this awkward situation will possibly be relatively common.[45] We would like to conclude by defending this claim, as it is likely to face resistance. Many authors have pointed out that the popular press oversimplifies the methods and results of neuroscientific studies.[46] The results are presented as more generalizable than they actually are, and it has been claimed that researchers are able to do things which, so far at least, they cannot do at all.[47] For instance, the use of neuroimaging to gather information about people's psychological traits is possible today but only to a very limited extent.[48] The popular press tends to create not only false hopes and overly optimistic scenarios but also unjustified worries.[49] Our assumption in this paper is compatible with these claims, and we stress rather than deny that many expectations concerning neuroscience innovations are, at present, mistaken. However, it seems relatively clear that the future can be different. It is likely that fMRI and similar technologies will provide all kinds of information about people's inner lives. In principle, fMRI lie detection and diagnostic neuroimaging, for instance, may also reveal something that *was not searched for* by

anyone.[50] Access to online databases involving private information about people's mental lives may result in worrisome effects. There are also likely to be *more and more application fields* of fMRI.[51] It follows that there will be more people whose brains will be scanned in the context of health care, research, employment, insurance, criminal justice, litigation, and so on. The overall consequence of all this, as we see it, is that knowledge about conflicting attitudes *will* probably be common, or at least considerably more common than it presently is. The question of exactly when and to what extent this will happen is beyond the scope of this paper. Philosophical reflection on the topic is justified in any case, and we would do well to prepare ourselves.[52]

Juha Räikkä

University of Turku

Saul Smilansky

University of Haifa

NOTES

1. Georges Rey has distinguished between avowed beliefs and central beliefs. "Toward a Computational Account of *Akrasia* and Self-Deception" in B.P. McLaughlin and A.O. Rorty, eds., *Perspectives on Self-Deception* (Berkeley: University of California Press, 1989), 264–95, esp. 280. Philip Pettit distinguishes behavioural beliefs and judgmental beliefs in "Practical Belief and Philosophical Theory," *Australasian Journal of Philosophy* 76 (1998), 15–33, esp. 18. The notion of judgment-sensitive attitudes is used for instance by Thomas Scanlon. *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 2000), 18–22. Judgment-sensitive attitudes are attitudes that, in ideally rational agents, are sensitive to reasons, such that these agents have them when, and only when, they judge there to be sufficient reason for them. John Martin Fischer and Mark Ravizza talk about reason-responsive "mechanisms" in *Responsibility and Control* (Cambridge: Cambridge University Press, 1998), 69–76.

2. We understand "psychological evidence" in the broad sense so that it includes evidence provided by brain sciences and cognitive neuroscience. Hence "neural evidence" is interpreted roughly as "psychological evidence."

3. Richard Moran writes about the "kind of alienation" that people face when they are aware of their beliefs merely on theoretical grounds (*Authority and Estrangement* [Princeton: Princeton University Press, 2001], 92). This paper has benefited considerably from Moran's book. The idea that one's beliefs are transparent to oneself has been recently criticized by Brie Gertler. See "Self-Knowledge and the Transparency of Belief" in A. Hatzimoysis, ed., *Self-Knowledge* (New York: Oxford University Press, 2011), 125–45.

4. There is a large amount of evidence that people are surprisingly often unaware of the considerations that influence their behaviour. See, e.g., Timothy D. Wilson, *Strangers to Ourselves* (Cambridge, MA: Harvard University Press, 2002). See also Eric Schwitzgebel, "The Unreliability of Naïve Introspection," *Philosophical Review* 117 (2008), 245–73. Schwitzgebel argues that we are "prone to gross error, even in favourable circumstances of extended reflection, about our ongoing emotional, visual, and cognitive phenomenology" (p. 259).

5. Cf. D.H. Mellor, "Conscious Belief," *Proceedings of the Aristotelian Society* 78 (1977–78) 87–101, esp. 97; Rey, "Toward a Computational Account of *Akrasia* and Self-Deception," 281; Richard Moran, "Problems of Sincerity," *Proceedings of the Aristotelian Society* 105 (2005), 341–61, esp. 357.

6. Gary Watson writes that feelings of irrational guilt can be merely an "acculturated attitude" that is compatible with a seemingly conflicting value judgment. See "Free Agency," *Journal of Philosophy* 72 (1975), 205–20, esp. 215.

7. Damian Stanley, Elisabeth Phelps and Mahzarin Banaji published an fMRI-based study that appeared to suggest that people may have unconscious racist biases and that it can be detected. However, the researchers warned against interpreting the results as indicating hidden racism. See "The Neural Basis of Implicit Attitudes," *Current Directions in Psychological Research* 17 (2008), 164–70.

8. The saying "A man's true secrets are more secret to himself than they are to others" derives from Paul Valery.

9. Our definition of alien attitudes is not directly based on the debate concerning how to distinguish those desires that a person identifies with from those that are alien to her. For a discussion of that issue, see, e.g., Harry Frankfurt, "Freedom of the Will and the Concept of a Person," published in his *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988); James Tracey Taylor, "The Myth of Objectively Alien Desires" in D.K. Chan, ed., *Moral Psychology Today* (UK: Springer, 2008), 109–22. What Frankfurt calls "alien desires" are not a type of "alien attitudes" in our sense.

10. Richard E. Nisbett and Timothy DeCamp Wilson, "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review* 84 (1977), 231–59, esp. 241–42. The stockings experiment is mentioned also by Rey in his "Toward a Computational Account of *Akrasia* and Self-Deception" (272).

11. An alternative interpretation is the belief that "the right-most stockings are better than the left-most stockings, because I noticed the right-most stockings later than the left-most stockings." Nisbett and Wilson (1977, 244) write that it is not obvious why the position effect occurs, but that it "is possible that subjects carried into the judgment task the consumer's habit of 'shopping around', holding off on choice of early-seen garments on the left in favour of later-seen garments on the right."

12. Nisbett and Wilson (1977, 247) point out that "if people were aware of position effects on their evaluation, they would attempt to overcome those effects."

13. Maureen Sie, "Moral Agency, Conscious Control, and Deliberative Awareness," *Inquiry* 52 (2009), 516–31, esp. 520.

14. It is often said that groundless optimism is important in daily routines, not only because it contributes to psychic health, but also because positive (but false) beliefs often help us in our undertakings (by enhancing our confidence in our own abilities, and the like.). This is why there are good prudential reasons to have false rather than true beliefs in certain circumstances. However, people cannot choose their beliefs "at will." Cf. Daniel

Goleman *Vital Lies, Simple Truths* (New York: Simon Schuster, 1985); Shelley E. Taylor, *Positive Illusions* (New York: Basic Books 1989). Saul Smilansky argued in *Free Will and Illusion* (Oxford: Oxford University Press, 2000) that we should continue to deceive ourselves on the free-will problem. For a discussion of the ethics of self-deception, see, e.g., Mike W. Martin, *Self-Deception and Morality* (Lawrence, KS: University Press of Kansas, 1986).

15. Here is another example. Before elections a responsible citizen may fail to judge that "my vote is important," but if he in fact believes so, then he has a morally desirable alien attitude, given that he is aware of his evidential belief and thinks (perhaps correctly) that having it is morally important, as it indicates his commitment to democracy. (Morally desirable alien attitudes can be true or false.)

16. Why would a person have a desirable alien attitude? Because of "twisted" self-deception, for instance: a person can be self-deceived into believing something she does not want to be true. Cf. Mele, *Self-Deception Unmasked* (Princeton: Princeton University Press, 2001), Ch. 5. Morally desirable alien attitudes do have a psychological function.

17. Christopher Peacocke, *Being Known* (Oxford University Press, Oxford 1999), 242–43. Peacocke defends the claim that we may have introspectively justified but nevertheless false second-order beliefs. Aaron Z. Zimmerman criticizes Peacocke's example in "Self-Knowledge," *Philosophy Compass* 3 (2008), 325–52.

18. The principle of belief attribution used here could be the following: "If the best explanation of S's non-verbal behavior includes attributing to S the belief that not-p, then do not attribute to S the belief that p." Cf. Steven D. Hales, "Self-Deception and Belief Attribution," *Synthese* 101 (1994), 273–89, esp. 287.

19. In Akeel Bilgrami's thought experiment a person has self-knowledge of his thoughts only from an external or "third personal perspective on himself." See *Self-Knowledge and Resentment* (Cambridge, MA: Harvard University Press, 2006), 190–91.

20. Obviously, being biased does not always mean that one has a biased *belief*. A person may have biased "impressions," and so on.

21. Neil Levy discusses "the faulty attitudes" and "the epistemic conditions on moral responsibility." See "The Good, the Bad, and the Blameworthy," *Journal of Ethics & Social Philosophy* 1 (2005), 2–16, esp. 11.

22. Cf. Moran, *Authority and Estrangement*, 63.

23. Levy's claim that judgment-sensitive attitudes "are responsive to reasons such that when no such reason is forthcoming, they tend to weaken and eventually extinguish" may be correct, but there are many exceptions to this tendency. See "The Good, the Bad, and the Blameworthy," 11.

24. Perhaps a person who manages to *internalize* his precautionary measures in the sense that he need no longer think of them nor intentionally apply them *has* managed to get rid of his biased belief. But this is not clear.

25. If she is unable to work wholeheartedly with people who have undergraduate degrees from countries other than her own, why does she not just quit? Because then most professors should quit? Because all things considered she is perhaps the best person for the job? Because she is irreplaceable? There may be, in other words, good reasons for her to remain on the job.

26. Robert Merrihew Adams, "Involuntary Sins," *The Philosophical Review* 94 (1985), 3–31, esp. 15.

27. Adams, "Involuntary Sins," 15. Adams (1985, 19) argues that a person's "evil beliefs are a part of who he is, morally, and make him a fitting object of reproach," no matter what is the origin of his evil beliefs.

28. Angela M. Smith thinks that what makes us responsible for our attitudes is that they are the kinds of states that "reflect and are in principle sensitive to our rational judgments." See "Responsibility for Attitudes: Activity and Passivity in Mental Life," *Ethics* 115 (2005), 236–71, esp. 271.

29. Cf. Moran's discussion on unmotivated desires in *Authority and Estrangement*, 115.

30. Cf. Moran, *Authority and Estrangement*, 67, 151.

31. Georges Rey argues in "Toward a Computational Account of *Akrasia* and Self-Deception" (1988, 282–83) that people can *choose* whether they identify with their "avowed attitudes" or with what he calls "central attitudes." Rey writes that just "as avowals themselves are caught up in social relations, so are the identifications we make with them." "On the one hand, one doesn't want to be overly burdened with the biases, superstitions, and stupidities that one may centrally believe despite one's better (avowed) judgment; but, on the other hand, one oughtn't to be swayed by now this, now that bit of explicit reasoning."

32. The CEO's "obligation" to think that "women are as able as men to act in responsible leadership positions in business enterprises" is a role-specific obligation. He has this obligation in virtue of being the CEO of a big company. In general it would be wrong to conclude that a person who has never considered the question of how good men and women are in business and therefore does not believe that "women are as able as men to act in responsible leadership positions in business enterprises" is a terribly bad person.

33. Rey describes the case of an "educated" neurotic who "might acquiesce to the nonconscious motives ascribed to him by his therapist." See "Toward a Computational Account of *Akrasia* and Self-Deception," 276.

34. A separate ethical question is how easily a person should trust belief-attributions that concern his or her own beliefs and attribute to himself or herself beliefs that conflict with his opinions.

35. When a person feels that he is the only person who could eradicate his alien attitude, he has in mind *natural* ways of dealing with the issue. Of course there are medical and technical means to eradicate people's beliefs, alien or not.

36. See Leon Festinger, *A Theory of Cognitive Dissonance* (Stanford: Stanford University Press, 1957).

37. Epistemic conservatism comes in many forms, but roughly speaking it is the doctrine that claims that we may have a justification for a belief in virtue of holding that belief. For a defence of epistemic conservatism, see. e.g., Kevin McCain, "The Virtues of Epistemic Conservatism," *Synthese* 164 (2008), 185–200. David Christensen has criticized epistemic conservatism in his "Conservatism in Epistemology," *Nous* 28 (1994), 69–89.

38. There is an air of self-deception if the CEO manages to change his opinion by interpreting "evidence" in the way that helps him to form a belief that he wishes to have. See Alfred Mele, *Self-Deception Unmasked*, 25–31.

39. Since Daniel Dyke's *The Mystery of Self-Deceiving* (London: Griffin and Mab, 1614), an enormous number of contributions have been written about the mystery and ethics of self-deception.

40. Moran, "Problems of Sincerity," 357.

41. Of course, the CEO's report about his state of mind is also an opinion of his, i.e., his opinion about his state of mind.

42. Someone might say that the CEO takes advantage of Moore's paradox and plays with two propositions: "I believe that women are as able as men to act in responsible leadership positions in business enterprises" and "women are not as able as men to act in

responsible leadership positions in business enterprises." But this does not appear to be the accurate picture, as the CEO's attitudes seem to be his (alien) evidential belief that "women are as able as men to act in responsible leadership positions in business enterprises" and his opinion that "it does not seem to me that women are as able as men to act in responsible leadership positions in business enterprises."

43. Obviously, if a person who has the evidential belief that "a conservative candidate is better than a liberal candidate" votes for the liberal candidate, then this voting should be taken into account when it is considered whether the belief "a conservative candidate is better than a liberal candidate" can still be attributed to the person.

44. A husband may wonder whether his wife *really* loves him, not whether his wife sincerely thinks that she loves him. Such a husband would like his wife to analyze her emotions from a third-person point of view.

45. The point here is that neuroscience will reveal contradictions in our attitudes, not that it will play a causal role in producing them.

46. See, e.g., Eric Racine, *Pragmatic Neuroethics* (Cambridge, MA: The MIT Press, 2010), Ch. 5.

47. Cf. Valtteri Arstila, "Brain Reading and the Popular Press," *Res Cogitans* 8 (2011), 4–24.

48. See Martha J. Farah, et al., "Brain Imaging and Brain Privacy: A Realistic Concern?," *Journal of Cognitive Neuroscience* 21 (2008), 119–27.

49. Cf. Arstila, "Brain Reading and the Popular Press," 4–5.

50. See, e.g., Margaret L. Eaton and Judy Illes, "Commercializing Cognitive Neurotechnology—The Ethical Terrain," *Nature Biotechnology* 25 (2007), 393–97. Valtteri Arstila and Franklin Scott argue that the idea that brain-imaging data could reveal something unintended (i.e., something that the researchers did not look for in the first place) holds for "structural characteristics of our cortex where abnormalities are often found in MRI scanning," but unintended findings are *unlikely* "in the cases where researchers aim at investigating more 'dynamic' states, such as thoughts, memories, and personality traits." See their "Brain Reading and Mental Privacy," *Trames* 15 (2011), 204–12, esp. 208.

51. Cf. Martha J. Farah, "Neuroethics: The Practical and the Philosophical," *Trends in Cognitive Sciences* 9 (2005), 34–40.

52. We would like to thank Luca Barlassina, Walter Glannon, Amihud Gilead, Iddo Landau, Ariel Meirav, Daniel Statman and Jukka Varelius for helpful comments.