

Mill's Proof and the Guise of the Good

Francesco Orsi*

Department of Philosophy, University of Tartu, Estonia

The guise of the good doctrine is the view that whatever we desire, we desire it under the guise of the good, i.e. it appears good to us in some way. In this paper I first clarify the role that the doctrine of the guise of the good plays in the first step of J. S. Mill's proof of the principle of utility (in which he shows that one's happiness is desirable as an end). Then I provide textual evidence in favour of ascribing the doctrine to Mill, arguing that he commits to it to the extent that he equates finding something pleasant and thinking it desirable. Finally I counter two potential sources of evidence against ascribing the guise of the good to Mill: apparent desires based on "fixed ideas", and those habitual desires which are no longer associated with finding their objects pleasant. I argue that "fixed ideas" do not feed actual desires, and that the habitual desires which seem to escape the guise of the good, even if not uncommon, have a secondary status as desires.

Keywords: John Stuart Mill; guise of the good; desire; utilitarianism; value theory.

1. Introduction

The doctrine of the guise of the good is the view that whatever we desire, we desire it under the guise of the good, i.e. it appears good to us in some way. In this paper I first clarify the role that the doctrine of the guise of the good plays in the first step of J. S. Mill's proof of the principle of utility (in which he shows that one's happiness is desirable as an end) (section 2). Then I provide textual evidence in favour of ascribing the doctrine to Mill, arguing that he commits to it to the extent that he equates finding something pleasant and thinking it desirable (section 3). Finally I counter two potential sources of evidence against ascribing the guise of the good to Mill: apparent desires based on "fixed ideas", and those habitual desires which are no longer associated with finding their objects pleasant. I argue that "fixed ideas" do not feed actual desires, and that the habitual desires which seem to escape the guise of the good, even if not uncommon, have a secondary status as desires (section 4). I conclude that we can ascribe to

*Email: francesco.orsi@ut.ee

Mill a coherent, even if not fully general, guise of the good doctrine.

2. GG and the first step of Mill's proof

Several scholars attribute to Mill a certain view about desire in order to defend, or at least make sense of, his (in)famous claim that “the sole evidence it is possible to produce that anything is desirable, is that people do actually desire it” (U 4.3).^{*} This claim is the overarching premise of Mill's proof of the principle of utility, in which he aims to show that “happiness is desirable, and the only thing desirable, as an end” (U 4.2). It is common to view Mill's proof as consisting of three steps:

- 1) showing that one's own happiness is one of the things desirable as an end (or for its own sake), as far as one is concerned (most of U 4.3);¹
- 2) showing that one's own happiness is the only thing desirable as an end, as far as one is concerned (U 4.4-11);
- 3) showing that the general happiness or happiness as such is the only thing desirable as an end (final lines in U 4.3).²

While steps 1 and 2 are concerned with establishing just what is desirable as an end or for its own sake, and whether there is more than one such thing, the third step answers the rather different question whether the intrinsic or final value of happiness should be regarded as agent-relative (I have special reasons to care about my happiness as such) or as agent-neutral (my reasons to care about my happiness are in principle also reasons to care about anyone else's happiness).³ In this sense, there is a division of labour between the first two steps and the third one. The first two steps are thus worth investigating in their own right, even if the proof is not complete without the third step.

The view of desire that many attribute to Mill can be formulated as a version of the guise of the good doctrine (GG from now on): whatever we desire, we desire it under the guise of the good, i.e. it appears good (or desirable) to us.⁴ How would exactly GG help Mill's proof? In this section I intend to lay out the role that GG should be seen as having in the first step. In doing this, I do not claim particular originality for my reconstruction—the point is rather to draw on the existing state of the art regarding the role of GG in Mill's proof, in order to formulate a single, coherent argument. In other words, while the guise of the good has been evoked in Mill scholarship for a while (though not necessarily under that name), no one has yet

(to my knowledge) collected the suggestions made (often independently) by different commentators to make them converge into one interpretive line. Moreover, these suggestions are typically made *en passant* with little effort to spell out precisely what Mill's argument would look like were he to avail himself of GG.

The idea generally shared by commentators on this point is that GG would support Mill's claim that what we desire is the sole evidence that something is desirable.⁵ But how would the argument exactly proceed? Here is my proposal. By GG, if we desire X, then X appears desirable to us. Since we do desire X, X appears desirable to us. But the way things appear to us is, normally, at least defeasible evidence that they are the way they appear. Call this *the appearance-as-evidence premise*. Therefore, that X appears desirable to us is defeasible evidence that X is desirable.⁶ Thus, insofar as we desire X, we have defeasible evidence for believing that X is desirable (or as Mill says, what we desire is evidence that something is desirable). Substitute "our own happiness" for "X", and add "as an end" as a qualifier for the type of desire we have for our own happiness, and you obtain the first step of Mill's proof: one's own happiness is one of the things desirable as an end.

The appearance-as-evidence premise is arguably what Mill tacitly appeals to in his controversial analogy between what is desired/what is desirable and what is seen/what is visible: "The only proof capable of being given that an object is visible, is that people actually see it. The only proof that a sound is audible, is that people hear it: and so of the other sources of our experience. In like manner, I apprehend, the sole evidence it is possible to produce that anything is desirable, is that people do actually desire it" (U 4.3). Geoffrey Sayre-McCord argues that Mill is drawing an analogy between "X is desired, i.e. X appears desirable, therefore (defeasibly) X is desirable" and, for example, "X is seen as red, i.e. X appears red, therefore (defeasibly) X is red" (2001, 339). In both cases, an inference is made from the way things appear.

There is, however, a textual problem here: the analogue of "desirable" for Mill is officially "visible", not "red" or any other visible property. So, if one were to apply the appearance-as-evidence premise to Mill's official analogy, the result would be rather odd: the analogy would have to be between "X is desired, i.e. X appears desirable, therefore (defeasibly) X is desirable" and "X is seen, i.e. X *appears visible*, therefore (defeasibly) X is visible." But it is dubious that for something to be seen is for it to appear visible.

This said, I believe the appearance-as-evidence premise (and with it the GG interpretation) can still be saved. Inferring that an object (or a property) is visible or audible is

tantamount to inferring that the object is real (or that it really has a certain property): if something is visible, it is not just that “it can be seen”—it really exists, or at least it exists *as a visible object*. Thus, moving from what we actually see or hear to what is visible or audible is nothing but moving from what appears to be real (where the appearance is visual or auditory) to what is, visibly or auditorily, real. Analogously, moving from what we desire to what is desirable is nothing but moving from what appears to be desirable to what is really desirable. In other words, the analogy is ultimately between, on the one hand, empirical appearances and empirical reality, and on the other hand, value appearances and value.

The role played by the guise of the good in the first step of Mill’s proof should now be clear. Without the assumption that desiring X entails X appearing desirable, the appearance-as-evidence premise would not be applicable, and therefore Mill would not be able to move from what is actually desired (for its own sake) to what is desirable (for its own sake). But we should note that GG continues to play a role in the second step of Mill’s proof as well: because otherwise Mill could not move from his finding that people *only* actually desire their own happiness as an end to the conclusion that *only* happiness is desirable as an end.

I will not defend this interpretation against others that *do not* ascribe GG to Mill. The point of this section was simply to organize the existing scholarly proposals into one coherent argument. I conclude this section by noting (but not addressing) possible liabilities of this strategy. First, the first step of the proof, thus understood, would not be conclusive: for all it says, there might well be counterevidence or defeaters which undermine the evidential status of the appearance that our own happiness is desirable for its own sake (in fact, in step 3 of the proof it turns out that our own happiness, *qua our own*, is not desirable for its own sake). Second, we have not vindicated Mill’s strong claim that what people actually desire is the “sole” evidence that something is desirable (if destroying trees for fun did not appear undesirable to anyone—if nobody had an aversion to it—would there not be *any other possible* evidence that it is undesirable?). I will next investigate the textual grounds for attributing GG to Mill, and in which form.

3. The evidence for GG in Mill

Though many scholars have shown why GG would be useful to Mill’s proof, there is no detailed discussion of whether Mill actually held or could have held GG. In this and the next section I take up this task.

To my knowledge, Mill never makes any claim as explicit as GG, nor does he discuss the doctrine anywhere. He gets close to GG when he says that, thanks to association with past pleasures, virtue “may be *felt a good in itself*, and *desired as such* with as great intensity as any other good” (U 4.7, my emphasis). Even this sentence, though, simply juxtaposes something being desired and something being “felt” as a good. Also, he might be implicitly accepting something like GG in his claim (in the very context of the first step of the proof) that “[i]f the end which the utilitarian doctrine proposes to itself were not, in theory and in practice, acknowledged to be an end, nothing could ever convince any person that it was so” (U 4.3). If “end” here means something *worth* pursuing or desiring, then Mill is saying that his proof moves from what we desire, i.e. acknowledge (in some way) to be desirable or worth desiring, to the utilitarian doctrine.

However, GG can be more securely attributed to Mill in a slightly roundabout way, which will bring out Mill’s particular take on the guise of the good. As the proof moves forward, Mill states a number of psychological equivalences, of which GG seems to be a straightforward implication. Here is the first equivalence:

[D]esiring a thing and finding it pleasant, aversion to it and thinking of it as painful, are phenomena entirely inseparable, or rather two parts of the same phenomenon; in strictness of language, two different modes of naming the same psychological fact. (U 4.10)

And here is the second equivalence:

[T]o think of an object as desirable (unless for the sake of its consequences), and to think of it as pleasant, are one and the same thing. (U 4.10)

A simple argument can now be constructed:

1. Desiring X for its own sake = finding X pleasant
2. Finding (“thinking”) X pleasant = thinking of X as desirable for its own sake
3. Therefore, desiring X for its own sake = thinking of X as desirable for its own sake

The conclusion obviously implies that whatever we desire (for its own sake) appears desirable (for its own sake) to us. The conclusion is itself a strong version of the guise of the good, whereby desires just *are* evaluative thoughts. It is crucial not to read too much into Mill’s talk of “thinking of X as desirable”: it certainly need not amount to a *judgment* that X is desirable.

Despite its vagueness, the notion of an “evaluative appearance” is probably the best way to describe the psychological facts here. For us to find X pleasant, and in turn for us to desire X, is for X to appear good to us. And for X to appear good to us is for X (or the thought of X) to *feel good* to us. Premise 2 is in fact confirmed by other texts, where Mill even treats “pleasant” and “desirable” as interchangeable:⁷

The idea, for example, of a given *desirable object*, will excite in different minds very different degrees of intensity of desire [...] the idea of *some particular pleasure* may excite in different persons, even independently of habit or education, very different strengths of desire. (*A System of Logic*, CW 8: 857-858, my emphasis)

[T]he doctrine of the causation of our volitions by motives, and of motives by the desirable objects offered to us, combined with our particular susceptibilities of desire... (ibid.: 844)

In the latter quote, “desirable” is interchangeable with “pleasant”, because Mill is summing up his theory of motives, where ideas and sensations of pleasure (and pain) play the central role, as was already suggested in premise 1 above.

However, precisely a closer look at his theory of motives reveals that premise 1 in our reconstruction above cannot be quite right. In his notes to his father’s *Analysis*, Mill writes:

I believe the fact to be that Desire is not Expectation, but is more than the idea of the pleasure desired, being, in truth, the initiatory stage of Will. In what we call Desire there is, I think, always included a positive stimulation to action; either to the definite course of action which would lead to our obtaining the pleasure, or to a general restlessness and vague seeking after it. The stimulation may fall short of actually producing action: even when it prompts to a definite act, it may be repressed by a stronger motive, or by knowledge that the pleasure is not within present reach, nor can be brought nearer to us by any present action of our own. Still, there is, I think, always, the sense of a tendency to action, in the direction of pursuit of the pleasure, though the tendency may be overpowered by an external or an internal restraint. So also, in aversion, there is always a tendency to action of the kind which repels or avoids the painful sensation. (CW 31: 215)

In this note, as well as in others (CW 31: 249-250), Mill argues that “there is in a desire something inherently distinct from either an idea or an expectation” of pleasure (CW 31: 250). If all desire amounts to is an idea of pleasure, we would not explain why unsatisfied desire is painful—the mere notion of an “unsatisfied idea” clearly does not help. Instead, an unsatisfied

desire is painful because, in addition to an idea or expectation of pleasure, there is a distinct element in desire—the positive stimulation to action, and its phenomenological concomitant, a “sense of effort”—which finds no relief when the object of desire is not attained.

Therefore, it seems that, despite the equivalence stated in *Utilitarianism*, on Mill’s considered view desiring for its own sake and finding pleasant cannot be identical. Still, it seems fair to attribute to Mill the claim that desiring X for its own sake at least *necessitates* finding X pleasant. It is not clear whether he would accept the converse as well: whether finding X pleasant necessitates desiring X for its own sake. Perhaps finding X pleasant or painful may fail to be followed by the active element (stimulation to action) discussed above, and so one may find X pleasant without thereby desiring X.

In any case, once we take into account Mill’s notes, we can still offer a reconstruction that commits Mill to a weaker (and possibly more plausible) version of the guise of the good (3* is weaker than 3 because it falls short of reducing desires to evaluative thoughts):

- 1.* If one desires X for its own sake, then one finds X pleasant
2. Finding (“thinking”) X pleasant = thinking of X as desirable for its own sake
- 3.* Therefore, if one desires X for its own sake, one thinks of X as desirable for its own sake.

Clearly now we can see that if Mill subscribes to GG, it is because he subscribes to what we may call an *evaluative conception of finding something pleasant*, as formulated in premise 2. The claim in premise 2 might of course be alternatively termed “an hedonic conception of thinking X desirable for its own sake”. While I see no reason to choose between these two labels, two qualifications about premise 2 are in order.

First, as remarked above, Mill is not equating finding X pleasant (and in turn desiring X) with *judging* that X is good for its own sake, but only with X appearing or being thought of as good or in a positive light. Mill’s own texts make room for a distinction between evaluative judgments and the different attitude that I have called evaluative appearance.⁸ For example, when we choose to pursue a pleasure L, which is *by our lights* a “lower” pleasure than the “higher” pleasure H, at that moment we must be finding (the prospect of) L more pleasant than (the prospect of) H, or we would not desire L more than H (note: it is not necessarily that we expect to gain more pleasure from pursuing L than from pursuing H, it is just that the idea of L strikes us more pleasantly than the idea of H). According to my reconstruction (and following

Mill's equivalences), this implies that, at that moment, somehow we do regard L as more desirable than H. But Mill makes it clear enough that our preference for L over H in this case goes against our own evaluative *judgment* (U 2.8): after all, we do think of L as a lower pleasure compared to H. So when Mill identifies finding pleasant and thinking desirable, with the latter he must have in mind a different evaluative attitude than evaluative judgment.⁹

Second, even if finding X pleasant is not the same as judging X to be desirable, it might seem too intellectualistic to identify pleasures with *any sort* of evaluative attitude. Can't you experience some pleasure and be, so to speak, completely evaluatively indifferent to it? First, to clarify Mill's view, it is important to point out that Mill is not equating any old state of pleasure (or pain) with thinking desirable (or undesirable) for its own sake. In the quotes reported above Mill refers to *finding X pleasant*, *thinking of X as painful*, and to the *idea* of a pleasure (or the idea of a pain). These are naturally interpreted as *hedonic attitudes* we have towards objects or states of affairs, rather than as hedonic sensations or purely experiential states. To this extent, they have the right "structure" to allow for the evaluative gloss that (according to this interpretation) Mill reads into them.¹⁰ Second, as suggested in the quote from the notes to his father's work, these hedonic attitudes must participate in the "initiator stage of the Will" by guiding desire and in turn action towards (or away from) certain objects. Even if Mill does not typically use the language of reasons for action, it seems that nowhere else but in these hedonic attitudes could the agent's reasons for action¹¹ be found within Mill's philosophy—the agent's answer to the question "why/for which reason did you do X?" must ultimately refer back to what the agent thinks of as pleasant or painful.¹² And for such an answer to have the rationalizing or "sense-making" power it is ordinarily supposed to have, it is plausibly necessary (though perhaps not always sufficient) to understand hedonic attitudes as evaluative in nature: they must point to something regarded as desirable, by the agent herself if by nobody else.¹³ In sum, it is plausible and not overly intellectualistic to regard Mill's "finding pleasant" (and its twin "thinking desirable") as involving a genuinely evaluative representation.

4. Evidence against GG in Mill

It might be thought that Mill's texts also contain evidence *against* attributing GG—and so against the scholarly plausibility of using GG to make sense of the first step of his proof. I will describe, and deflect, two such challenges. The first is the idea that GG fails to capture some apparent desires. Mill writes (in part to criticize his father's account of voluntary action as too narrow):

There are cases in which a vivid imagination of a painful fact, seems really to produce the action which realizes the fact. Persons looking over a precipice are said to be sometimes seized with a strong impulse to throw themselves down. Persons who have extreme horror of a crime, if circumstances make the idea of committing it vividly present to their mind, have been known, from the mere intensity of their horror, to commit the crime without any assignable motive; and have been unable to give any account of why they committed it, except that the thought struck them, that the devil tempted them, and the like. This is the case of what is sometimes called a fixed idea; which has a sort of fascinating influence, and makes people seek what they fear or detest, instead of shunning it [...] This peculiar case obliges us to acknowledge the coexistence of two different modes in which action may be excited. There is the normal agency of the ideas of a pleasure and a pain, the one determining an action towards the pleasure, the other an action away from the pain; and there is the general power of an extremely strong association of any kind, to make the action follow the idea. (CW 31: 250-253)

Apparent desires based on a “fixed idea” seem to be a counterexample to GG: the object of desire (e.g. throwing yourself off the precipice) is not presented *in any way* as desirable, but as undesirable. And Mill’s talk of “two different modes in which action may be excited” does suggest that fixed idea explanations may be a genuine alternative to the ordinary explanation where we desire (and pursue) what appears desirable, and we are averse to (and avoid) what appears undesirable. So, at best, Mill would only partially subscribe to GG. But that is tantamount to admitting that he cannot rely on GG for his proof; after all, some people in some circumstances would seem to desire their own unhappiness as an end.

I think there are two complementary answers to this challenge. The first is that, as Mill explains, fixed ideas gain control by overpowering the “specific tendency of a painful association to repel action” (ibid., 253). In other words, even in these cases we *are* still averse to what appears undesirable to us, though the fixed idea exerts a greater attractive force on us. The second answer is that Mill never describes the attractive force of a fixed idea in terms of a desire or a motive, but rather in terms of “impulse”, “fascinating influence”, and the absence of any “assignable motive”. So, while there are indeed two ways a voluntary action can be produced, it seems that the ordinary way involves a desire (and thus GG), while the one based on fixed ideas does not even involve a desire. Fixed ideas rather seem to *bypass* and *interfere with* the agent’s desires and aversions. So fixed idea explanations should not be regarded as an exception to GG.¹⁴

A second challenge to my reconstruction starts from Mill's claim that the motives which are the result of habit or of a "confirmed character" are not, or at least need not be, based on thinking or finding something pleasant: "a person of confirmed virtue, or any other person whose purposes are fixed, carries out his purposes without any thought of the pleasure he has in contemplating them, or expects to derive from their fulfilment" (U 4.11, and cp. *A System of Logic*, CW 8: 842ff.). Given the identity suggested above between finding something pleasant and thinking of it as desirable, it seems to follow that the person of confirmed virtue does not (or need not) think of pursuing virtue as desirable. But this seems an absurd implication. When virtue (or anything else) becomes a habit, we do not thereby cease to regard it as valuable—if anything, we are even more assured of its value. Since the absurd implication is the result of combining (a) the GG view I have attributed to Mill, with (b) Mill's claim about habitual motives, it seems I should not attribute him the GG view whereby finding pleasant and thinking desirable are one and the same thing (despite, admittedly, the equivalence stated by Mill in U 4.10).

However, this is a challenge that vanishes upon a closer reading of that particular passage: Mill says that the person of confirmed virtue actually *has* pleasure in contemplating her purposes, and even expects pleasure from their fulfilment, but she puts no thought to any of this when deciding to carry out her purposes. Far from asserting that habitual desires and motives float free of hedonic attitudes—and so, in turn, float free of the guise of the good—Mill is affirming here that very idea, while at the same time making what I take to be a point about the agent's reasons for action. With the onset of habits, many things we did exclusively *for* the pleasure of doing them or for the pleasures thereby causally produced (or for avoidance of pain) are now no longer (exclusively or at all) done for such hedonistic reasons. For those who become persons of "confirmed virtue", their own pleasure may disappear from the range of reasons in the light of which they pursue virtue. Similar things may be said for "confirmed" lovers of art or sports. But we do not thereby cease to find these activities pleasant (or their absence painful), and therefore we do not thereby cease to see them as desirable.¹⁵

Still, a harder nut to crack might be this passage from a related discussion in the *System of Logic*: "Although, from some change in us or in our circumstances, we have ceased to find any pleasure in the action, or perhaps to anticipate any pleasure as the consequence of it, we still continue to *desire* the action, and consequently to do it" (CW 8: 843, italics mine). In this quote, Mill seems to be considering the case of someone who *no longer* finds an activity pleasant (or its absence painful)—and so, on my reading of Mill's GG, would seem to no longer

find it desirable—and yet continues to desire it. Now this reads like a significant counterexample to GG—the objects of these desires are no longer presented as desirable. So it is a reason against attributing GG to Mill, if Mill’s GG is to depend on the finding pleasant/finding desirable connection.

One reaction could be to play around with what Mill is and is not saying here. Maybe these are cases of mismatch between states of enjoyment and what I have called hedonic attitudes. When doing sports has become an entirely joyless affair, we have no enjoyment actually produced or even expected from doing sports. Simply put, we no longer enjoy doing sports, nor do we expect that we will enjoy some of its consequences. However, the *thought* of doing sports has not necessarily ceased to be pleasant (or alternatively, the thought of not doing sports has not ceased to be painful): a positive hedonic attitude towards doing sports (or a negative hedonic attitude towards not doing sports) is still there. In other words, I may still find pleasure in the thought without finding pleasure in (getting pleasure from) the action. Here is another example. I may have ceased to enjoy my job and even its alleged benefits now leave me indifferent. I might even be aware of the enjoyments that await me if I were to quit. But I keep at the job and perform the tasks it requires, because “I can’t let go” or, in other words, the thought of quitting still strikes me painfully—I have a negative hedonic attitude towards quitting the job. And it is these positive or negative hedonic attitudes—and thereby a positive or negative evaluation—that can still feed our desires for doing sports or keeping the job, despite our lack of enjoyment from the activity itself or its consequences.¹⁶

But perhaps this is splitting one hair too many. A second reaction is to bite the bullet and concede that, when an activity has no longer *any association whatsoever* with our own pleasure, the activity can still be the object of our desire, without *thereby* being presented as good. When doing sports becomes a joyless affair, and even the thought of it is no longer pleasant, then we no longer positively evaluate doing sports *in* or *as part of* desiring doing sports. In other words, Mill would hold a restricted GG view applied only to some, but not all desires. Is biting this bullet a serious problem for my interpretation?

Let me explain why I believe this bullet is rather harmless. We should ask what feeds these pleasureless desires. Mill offers the following explanation: “in the case of an habitual purpose, instead of willing the thing because we desire it, we often desire it only because we will it” (U 4.11).¹⁷ The pleasureless desires that would escape Mill’s GG, though far from uncommon, seem to have a secondary status *as desires*. While pleasure-based desires play an active role as the initiatory stages of the will, pleasureless desires are in Mill’s own words the result of willing. Given Mill’s psychological associationism, pleasureless desires certainly bear

an historical relation with desires based on pleasure and pain (we once did enjoy doing sports or its benefits!). However, by habit and repetition (or “from some change in us or in our circumstances”), the relationship of will and desire can be reversed. In fact, it is tempting to read the “because” in “we often desire it only because we will it” as indicating a mere logical relation. Pleasureless desires can be ascribed to people when they act purely out of habit, but all the explanatory work is actually done by willing and habitual dispositions. In other words, when desires become pleasureless, they are only a logically necessary condition for action to take place, but are not “necessary either as a contributing influence, or as a causal condition” for action (Nagel 1970, 30). Following Bond (1983, 12), we could call such desires “logical shadows” of other, psychologically efficacious mental states. If the desires that fall short of Mill’s GG are only logical shadows, it seems perfectly unobjectionable to attribute to Mill a GG view that fails to capture *them*.

In this way I have deflected the two challenges to ascribing GG to Mill: actions based on a “fixed idea” are not cases of actions based on a desire; fixed or habitual purposes either leave room for hedonic attitudes (and so for Mill’s GG) or, when they include entirely pleasureless desires, these desires have a secondary status and one can be excused for not extending GG to them. There admittedly remains an open issue. Intuitively, an action which is desired and performed “only because we will it” may still be done for a motivating reason—in the light of a consideration the agent sees as favouring her action, whether or not she consciously deliberates about it. But if in these cases hedonic attitudes are not in the picture, the question arises, what the agent’s reasons for these actions could ever be (on Mill’s view). Is it the agent’s willing itself? Or perhaps the agent’s reasons are features of the action itself, it is just that willing something plays an analogous role to finding something pleasant—willing is the way in which certain actions appear good to us, and hence become objects of desire. If the latter is correct, then GG could be extended also to pleasureless desires. All desires present their objects as good—some do as a matter of finding something pleasant, others only as a matter of willing an action.¹⁸ But I am afraid Mill’s texts can hardly provide us with an answer here.

5. Conclusion

According to many commentators, GG could be used by Mill in supporting the first, controversial, step of his proof of the principle of utility. In my reconstruction, this step shows that we have at least some reason to believe that our happiness is desirable as an end, because

we desire it as an end, desiring our own happiness involves our own happiness appearing desirable (an application of GG), and if something appears desirable, then there is some reason to believe it is desirable. As I have argued in the rest of the paper, luckily, employing GG in the proof is not only required by interpretive charity, but it appears that Mill did commit himself to the doctrine, via the identification of my thinking X desirable (X appearing desirable to me) and my finding X pleasant. More work remains to be done regarding the nature and role of what I have called hedonic attitudes, as well as a comprehensive analysis of Mill's theory of reasons for action.¹⁹

Notes on contributor

Francesco Orsi is Associate Professor in Practical Philosophy at the University of Tartu. He has mainly worked in value theory, in particular on the fitting attitude account of value, and on the guise of the good.

References

- Alvarez, Maria. 2017. "Reasons for Action: Justification, Motivation, Explanation." *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2017/entries/reasons-just-vs-expl/>.
- Bond, E. J. 1983. *Reason and Value*. Cambridge: Cambridge University Press.
- Brink, David O. 2013. *Mill's Progressive Principles*. Oxford: Oxford University Press.
- Crisp, Roger. 1997. *Routledge Philosophy Guidebook to Mill on Utilitarianism*. London: Routledge.
- CW = J. S. Mill, *Collected Works of John Stuart Mill*, 33 volumes. Edited by John M. Robson. Toronto: University of Toronto Press, 1965–91.
- Macleod, Christopher. 2013. "Was Mill a non-cognitivist?" *Southern Journal of Philosophy* 51 (2): 206-223.
- Macleod, Christopher. 2020. "John Stuart Mill." *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2020/entries/mill/>.
- Mill, James. 1869. *Analysis of the Phenomena of the Human Mind*, 2 vols., 2nd ed., London: Longman, Green, Reader & Dyer.
- Miller, Dale E. 1998. "Internal Sanctions in Mill's Moral Psychology." *Utilitas* 10 (1): 68–81.
- Miller, Dale E. 2004. "On Millgram on Mill." *Utilitas* 16 (1): 96-108.
- Miller, Dale E. 2010. *J. S. Mill: Moral, Social and Political Thought*. Cambridge: Polity.

- Nagel, Thomas. 1970). *The Possibility of Altruism*. Oxford: Clarendon Press.
- Oddie, Graham. 2005. *Value, Reality, and Desire*. Oxford: Clarendon Press.
- Orsi, Francesco. 2015. “The Guise of the Good.” *Philosophy Compass* 10 (10): 714–724.
- Ryan, Alan. 1990. *The Philosophy of John Stuart Mill*. 2nd ed. London: Macmillan.
- Sayre-McCord, Geoffrey. 2001. “Mill’s “proof” of the principle of utility: A more than half-hearted defense.” *Social Philosophy and Policy* 18 (2): 330-360.
- Setiya, Kieran. 2010. “Sympathy for the devil”. In *Desire, Practical Reason, and the Good*, edited by Sergio Tenenbaum, 82-110. Oxford: Oxford University Press.
- Stocker, Michael. 1969. “Mill on desire and desirability.” *Journal of the History of Philosophy* 7 (2): 199-201.
- Stocker, Michael. 1979. “Desiring the bad: An essay in moral psychology.” *Journal of Philosophy* 76 (12): 738-753.
- Tenenbaum, Sergio. 2007. *Appearances of the Good: An Essay on the Nature of Practical Reason*. Cambridge: Cambridge University Press.
- Tenenbaum, Sergio. 2013. “Guise of the Good.” In *The International Encyclopedia of Ethics*, edited by Hugh LaFollette. Hoboken: Wiley-Blackwell.
- U = J. S. Mill (1861), *Utilitarianism* [CW X]. References by chapter and paragraph number.
- Zuk, Peter. 2018. “Mill’s Metaethical Non-cognitivism.” *Utilitas* 30 (3): 271-293.

Notes

* All references to *Utilitarianism* are to chapter and paragraph (as appears in the edition listed in bibliography).

¹ In what follows I deliberately use “intrinsic”, “final”, “as an end”, “for its own sake” as interchangeable (whether applied to values or to desires). This is because Mill himself makes no finer-grained distinctions among these expressions.

² This is e.g. Crisp’s reconstruction (1997, 72). For ease of exposition I have inverted the textual order of steps 2 and 3. I should add that this standard reading of Mill’s proof stands in contrast to Zuk (2018), who argues that in steps 1 and 2 Mill is intent on proving only the descriptive claim that we can desire only our own happiness as an end, thus using “desirable” in the purely psychological sense of “capable of being desired”—and leaving it to the reader to make the inference from this descriptive claim to “only happiness is good as an end”. On Zuk’s reading, Mill would not need anything like GG for his proof.

³ There might be other ways of describing the third step than the agent-relative/agent-neutral distinction. The point is simply to highlight the difference in focus from the first two steps.

⁴ The doctrine goes back at least to Aristotle, was given a standard formulation by Aquinas, and has made several comebacks in the history of philosophy, as documented in this special issue. For an introduction to the current debate on GG, see Tenenbaum (2013) and Orsi (2015).

⁵ Here is a representative sample. Sayre-McCord (2001, 339-340): “[A] person who desires *x* is *ipso facto* a person who sees *x* as desirable. Desiring something is, for Mill, a matter of seeing it under the guise of the good”. Crisp (1997, 75-77): “[Mill’s] argument requires that we *recognize* the object of our desire as good” (77). Miller (2004, 103): “Our desires are evidence of what is desirable because we desire that which appears to us to be desirable”, and also Miller (2010, 44-45). Stocker’s reconstruction (1969) logically entails the claim that for Mill we desire all and only the things which we believe desirable.

⁶ It should be plain that the evidence provided by appearances is strictly indexed to each subject. If your desire for your happiness gave *me* evidence that *your* happiness is desirable, Mill would not need the third step of his proof.

⁷ Compare James Mill: “The idea of a pleasure is the idea of something as good to have” (*An Analysis of the Phenomena of the Human Mind*, 2nd ed., vol.2, 151). J. S. Mill would seem to agree, if only because in his rather critical notes to this part of his father’s work, he doesn’t comment on this particular claim. These equivalences

do raise the question whether “desirable” for Mill is a genuinely evaluative or a merely descriptive term, and thus whether “thinking desirable” can be understood as a genuinely evaluative representation in line with GG. I tackle this question in the text (end of section 3). But, in general, if “desirable” did not stand for an evaluative concept, then in the first step of his proof Mill would not have shown anything more than that our happiness is an object of pleasure (or an object of desire) for us.

⁸ The idea that desires are, or necessarily involve, evaluative appearances which may conflict with, and possibly win over, our evaluative judgments is common among many proponents of GG. See Tenenbaum (2007) and Oddie (2005).

⁹ I take this proposal to be compatible with various metaethical interpretations of Mill’s view of evaluative and moral judgment, whether non-cognitivist (e.g. Ryan 1990, 193, Zuk 2018) or cognitivist ones (e.g. Brink 2013, see also Macleod 2013).

¹⁰ Aren’t Mill’s hedonic attitudes simply beliefs about expected pleasure or pain? I think not. Hedonic attitudes *are* pleasures (when positive) or pains (when negative) directed at certain objects or actions, not beliefs. Sometimes they are the result of an expectation of pleasure or pain, but like I said in the case of choosing a lower over a higher pleasure, it might simply be that the thought of the object strikes us pleasantly or painfully. As Mill remarks, those who desire virtue for its own sake do so “either because the consciousness of [virtue] is a pleasure, or because the consciousness of being without it is a pain, or for both reasons united” (U 4.8).

¹¹ By “the agent’s reasons for action” I mean what are usually called her “motivating reasons”—the considerations in the light of which she does what she does, and which she is in a position to offer as an answer to “why did you do that?”. These contrast with “explanatory reasons” (the facts or psychological factors which causally explain one’s actions) as well as with “normative reasons” (the considerations that speak in favour of an action). See Alvarez (2017).

¹² It is hard to say whether for Mill the agent’s reason is to be identified (a) with the agent’s pleasure or pain (“...because X gives me pleasure/prevents pain”), or (b) with the hedonic attitude itself (“...because I find X pleasant”), or again (c) with some features of an action, whatever they may be, which are as it were “highlighted” for the agent by their connection to her hedonic attitudes (as in: “I would not do X if I didn’t find X pleasant/X’s absence painful, but I don’t necessarily do X *for* my pleasure”). Given Mill’s psychological associationism, it seems that people’s reasons can develop from the first, purely hedonistic mode to the third one. While we at first want to act virtuously for the sake of some reward, we can in the long run want (and so find pleasure in) virtue for its own sake. In any case I take GG to be compatible with the first and the third option. The second option strikes me as less plausible in light of GG: what appears good to me about an object X is not, normally at least, the very fact that X appears good to me. Since (on GG) the agent’s reasons are the facts or features which appear good to the agent, and in Mill X appearing good and X being found pleasant are the same thing, it would be odd to identify the agent’s reasons with the hedonic attitudes themselves.

¹³ This is not the place to defend the general philosophical view that motivation or motivating reasons require evaluation. It has received serious criticism (Stocker 1979, Setiya 2010). The aim is just to show that Mill’s equation between finding pleasant (and in turn desiring) and thinking as desirable fits within the strand of thinking about reasons for action that is congenial to GG.

¹⁴ I thus agree with Miller 1998, 76 fn. 25. In terms of reasons, we can say that actions done under the influence of a fixed idea are explained by merely explanatory reasons, but not by “motivating reasons” or “the agent’s reasons” (see fn.12 above). Also note that the resulting action is still voluntary, to the extent that the agent could have resisted the influence of the fixed idea.

¹⁵ In fact, the second step of Mill’s proof (“only happiness is desirable for its own sake”) is based on the very idea that, since the person of confirmed virtue does find virtue desirable for its own sake, she thereby finds it pleasant. And if she finds it pleasant, then virtue is a part of her happiness, rather than something separate from her happiness. For her to desire virtue is for her to desire a part of her happiness. So virtue, even when it does appear intrinsically good, does not stand as a candidate for an intrinsic value different from happiness. (If we could desire virtue for its own sake *and* as separate from our happiness, then the appearance-as-evidence premise would support the belief that virtue is desirable for its own sake and not as a part of happiness. But for Mill it is not possible to desire virtue for its own sake *and* as separate from your happiness—you can think you do, but you would be confused. *Virtue as object of an intrinsic desire* is part of happiness.)

¹⁶ See previous footnote for the importance of this in the second step of Mill’s proof. Perhaps happiness can be constituted even by some activities that provide no actual enjoyment, if the mere thought of them is otherwise pleasant to us. Exercising virtue may provide no enjoyment, but it could still be true that we desire it for its own sake (and it is part of our happiness) “either because the consciousness of it is a pleasure, or because the consciousness of being without it is a pain, or for both reasons united” (U 4.8).

¹⁷ Note the word “often”. Certainly not all habitual desires are pleasureless!

¹⁸ I should note that, if merely willing an action sometimes explains why it appears good to me, and in turn why I desire to do it, some may take this case as a trivial application of GG, because value appearances (the good or at least the apparent good) would not be leading the will, but rather the other way around.

¹⁹ I wish to thank two anonymous referees for this journal for their helpful comments, and audiences in Lund and Tartu, who gave me feedback at various stages. This article was supported by grants PUT1630 (Estonian Research Council) and SHVHV16145T (TK145) (SA Archimedes).