Taking the Morality Out of Happiness

Markus Kneer
University of Zurich

Daniel M. Haybron
Saint Louis University

Draft: January 20, 2023 (*please do not cite without permission*)

Abstract: In an important and widely discussed series of studies, Jonathan Phillips and colleagues have suggested that the ordinary concept of happiness has a substantial moral component. For instance, two persons who enjoy the same extent of positive emotions and are equally satisfied with their lives are judged as happy to different degrees if one is less moral than the other. Considering that the relation between morality and happiness or self-interest has been one of the central questions of moral philosophy since at least Plato, such a result would be of considerable philosophical interest. On closer examination of the original research and new studies, we suggest that the data point to a different conclusion: in the dominant folk understanding of happiness, morality has no fundamental role. Findings seeming to indicate a moralized concept are better explained, we suggest, by folk theories on which extreme moral turpitude indicates that an individual suffers from psychological dysfunction.

1. Introduction

In an important and widely discussed series of studies, Jonathan Phillips and colleagues (2017) have suggested that the ordinary concept of happiness has a substantial moral component.[1] For instance, two persons who enjoy the same extent of positive emotions and are equally satisfied with their lives are judged as happy to different degrees if one is less moral than the other. Considering that the relation between morality and happiness or self-interest has been one of the central questions of moral philosophy since at least Plato, such a result would be of considerable philosophical interest: It suggests that the views of the ancient eudaimonists and those following their example are not in fact contrary to ordinary sensibilities, as has been widely supposed.[2] While lay opinion has limited standing among some philosophers, philosophical ethics nonetheless tends to rest heavily on intuitions, for example as evidence of the views that might eventuate from a process of reflective equilibrium. As others have argued, lay intuitions can be an important source of evidence for where reflection might realistically lead us.[3] If commonsense supports the idea that happiness requires virtue, then eudaimonists may not be facing the uphill battle many have assumed.

[1] See also Phillips et al. (2011), Newman et al. (2014), and Phillips et al. (2014). For further recent work on the folk concept of happiness, see Díaz & Reuter (2021), Kneer & Haybron (2019), Prinzing & Fredrickson (2022), Prinzing et al. (2022), Reuter et al. (2022), Yang et al. (2021). For interesting, related, work on the notion of a meaningful life, see Fuhrer & Cova (2022).

[2] For a representative example, see Sumner (1996).

[3] For discussion, see *inter alia* Ludwig (2007), Kauppinen (2007), Williamson (2007), Liao (2008), Horvath (2013), Horvath & Wiegmann (2016, 2022), Nado (2015, 2016), Alexander (2012), Machery (2011, 2017), Weinberg et al. (2010, 2012), Kneer et al. (2021), Buckwalter (2022).

Here we examine the evidence purported to favor a moralized understanding of the lay concept of happiness, along with the results of new studies, and suggest that the data point to a different conclusion: folk thinking about happiness appears to be divided, with one large strain of thought indicating a dominantly or purely psychological concept of happiness, or alternatively a conception of well-being, that has no clear moral aspect. Findings seeming to indicate a moralized concept are better explained, we suggest, by folk theories of happiness on which extreme moral turpitude constitutes strong evidence that an individual suffers from inner turmoil or other psychological dysfunction. Participants appear simply to refuse to accept stipulations to the contrary, or make their own inferences about psychological matters not covered by the experimental prompt. At this juncture, the balance of evidence suggests that morality forms no part of the concept of happiness employed by many people, though there is sufficient interpretive diversity that in a substantial proportion of ordinary uses it may play a role.[4]

The paper proceeds as follows: We will briefly outline the philosophical stakes of the question at hand, to clarify why these sorts of studies might be important (section 2). We'll then discuss and further analyze some of the findings reported by Phillip et al. (2017) in sections 3 and 4, and find that the folk concept/conception of happiness might be considerably less sensitive to morality than they take it to be. In section 5 we report a novel experiment, which provides further support for the idea that the dominant concept/conception of happiness is morality-insensitive. Section 6 concludes.

2. The philosophical stakes

The idea that happiness generally bears a positive relationship with moral behavior is not especially controversial, and indeed is something of a truism—"honesty is the best policy" and so forth. Virtue is not easily operationalized, but there appears to be a consensus among well-being researchers that the moral tend to be happier than the immoral.[5] But that truism involves a much weaker claim than the view at hand: that morality is part of the very *concept* of happiness, or that happiness is constituted at least in part by moral virtue. If that is the case, then being moral is part of what it *is* to be happy. The fully happy immoralist is not just ruled out by the vagaries of human psychology; such a person is inconceivable.[6] This is a striking and controversial claim, but it is at least in the vicinity of distinguished philosophical views like Aristotle's.

Yet it is less obvious than it may seem what the philosophical stakes are here. Importantly, there is no major philosophical view on which morality is standardly assumed to be part of the *concept* of happiness. The venerable notion that happiness consists at least partly in virtue is often, perhaps usually, taken to be a substantive ethical thesis and not a merely conceptual claim.[7] When Aristotle identified *eudaimonia* with a life of virtuous activity, he did not seem merely to be stating a definition. Rather, he was saying something of substance about what is good for us or benefits us, or how we ought to live. Just what manner of substance that sort of claim involves is pretty much the chief subject matter of metaethics, for while it is no great mystery how there might be conceptual truths—we can stipulate them—the idea that there are truths about what's good for us or what we ought to do is much harder to explain. Such truths can seem to be deeply mysterious—"queer," as Mackie (1990) put it. We need not concern ourselves with

[4] On interpretive diversity, see Nichols & Ulatowski (2007).

[5] E.g., Kesebir and Diener (2014), Ricard (2015), Keltner (2009), and Tiberius (2008, 2015).

[6] Or at least is impossible in a very strong sense, e.g. metaphysically impossible.

[7] Darwall (2002). The conceptual reading is not uncommon, however. See, e.g., Foot (2001), Toner (2006).

the resolution of such puzzles here, as the point is just to illustrate that philosophical claims about the relation between morality and happiness are not plainly conceptual.

This is not a serious problem for Phillips' et al.'s thesis, since they could just as well frame their findings as showing something about the folk *conception* of happiness, leaving it open whether this involves conceptual or other sorts of claims. But it is important to understand how the idea that happiness consists partly in morality might bear on the philosophical debates: happiness would likely have to be equivalent to a certain sort of value, more often framed as well-being, welfare, *eudaimonia*, or flourishing, or in more technical terms, prudential value.[8] Overwhelmingly, the philosophical debates over the role of morality in "happiness" have concerned questions of self-interest: whether morality necessarily benefits us, or whether one can profit from immorality. This has indeed been one of the central questions of philosophical ethics since at least Plato, among other things helping to animate the eudaimonist tradition in ancient and medieval ethics, which was largely concerned to defend an affirmative answer.[9] We'll refer to this view, that well-being consists at least partly in virtue, as *perfectionism* about well-being.[10] Modern moral philosophy has tended to take the opposite view, that the wicked can thrive, in great part because this is taken to be intuitively plausible if not obvious.[11] So evidence that the folk agree with the likes of Plato, Aristotle, and Aquinas that immorality is necessarily bad for us would be of considerable philosophical interest, suggesting that many philosophers' intuitions on the matter are not representative. Phillips' et al.'s results most plainly bear on philosophical debates, then, if the concept of happiness is also a concept of prudential value: the concept of well-being, more or less.

An alternative possibility is that HAPPINESS is a hybrid concept conjoining moral and nonmoral elements, for instance denoting the satisfaction of a good person. This appears to be the hypothesis favored by Phillips et al., as their results cluster with a range of findings from the experimental philosophy literature in which putatively descriptive concepts, like those of intentionality, knowledge, and causation, have proven sensitive to moral considerations. The suggestion is that these are "dual character" concepts, with both descriptive and evaluative components.[12] In the case of happiness, the thought appears to be that happiness is partly a matter of psychological states like being satisfied with one's life or having a positive emotional state, and partly a matter of being morally good.

It is not clear, however, what the philosophical import would be of the discovery that HAPPINESS is a dual character concept of that sort. To be sure, any finding about the contours of the ordinary concept of happiness is liable to have philosophical interest, given that

[8] There is some question about the precise equivalence of these terms, but theories bearing any of these levels are standardly taken to be competing accounts of some common subject matter. For discussion of the concepts of happiness and well-being in philosophical work, see e.g. Badhwar (2015), Besser (2021), Heathwood (2021), Raibley (2012), and Rossi and Tappolet (2016).

[9] See, e.g., Annas (1993). For a review of work on the virtue-well-being connection, see Baril (2015).

[10] Most views of this sort do exhibit the classical perfectionist schema involving a nature-fulfillment ideal of well-being such as Aristotle's, but we use the term broadly to cover any theory that takes virtue, excellence, or morality to be a basic element of well-being.

[11] E.g., Sumner (1996).

[12] Phillips et al. (2017). On dual character concepts, see, e.g., Knobe et al. (2013) and, for a review, Reuter (2019). For discussion regarding intentionality see e.g. Knobe (2010) and Alicke & Rose (2010), for causation see e.g. Sytsma (2019). Whether these findings constitute evidence for the hypotheses that many apparently descriptive concepts central to our daily lives are of dual character can be debated: It might well be that morality easily *biases* judgments of this sort. Morally charged application of apparently descriptive concepts is easily misinterpreted as evidence in favor of the dual character of such concepts.

‘happiness’ and cognates are such central terms in our practical vocabularies. In this case, HAPPINESS being a dual character concept would mean the most if not all philosophical accounts of the concept are false, at least as theories of the ordinary concept. That is not of course a trivial result. But why employ a concept like this, on which to ascribe happiness is to attribute both a positive mental state and moral virtue to an individual? Is this a useful concept, or a grue-some conjunction of disparate kinds?[13] Unless there is some point to using such a concept, the proper response might be to revise or even discard the concept, and indeed it is commonly acknowledged that revision is often necessary in philosophical work involving folk concepts like that of happiness.[14] (A natural suggestion along these lines would be to excise the moral appendage from the concept, focusing attention purely on the descriptive psychological component.) Even if the concept has *some* utility—perhaps there are situations where one might wish to assess a person’s hedonic state and moral character simultaneously—it may be that revised variants of the concept focusing on more natural-seeming kinds would better serve the purposes that animate everyday and philosophical interest in questions of happiness.

Consider, for instance, what happens to *comparisons* of happiness on such an understanding: if A is happier than B, does that mean A is morally better than B, feels better, or both? It is questionable how informative such a comparison would be. Problems like this commonly afflicted philosophical work on happiness in the days of linguistic philosophy, where theories of happiness tended to incorporate whatever features struck the theorist as intuitively plausible, without regard to questions about their theoretical or practical rationale. As a result it was unclear what of substance might be at stake in such theorizing, likely contributing to the marginal status of happiness as a philosophical topic until recent decades.[15]

Our point here is not to argue that HAPPINESS cannot be, or is not usefully understood as, a dual character concept with moral and psychological elements. It is simply to highlight that the philosophical and practical stakes in studies about the role of morality in ordinary happiness ascriptions are clearest if HAPPINESS is taken to be roughly equivalent to WELL-BEING. In that case, the Phillips et al. findings may be quite significant for an ancient, central debate about the relationship between morality and self-interest: Plato, Aristotle and many others’ claims that morality necessarily benefits us may not be counterintuitive after all, as they appear to reflect ordinary thinking about happiness. If, instead, the studies are taken to show that HAPPINESS is a dual character concept, or takes some other form not widely recognized in the philosophical literature, then their practical and philosophical upshot is far less apparent. Perhaps the folk concept just needs to go in for repairs.

Another feature of the philosophical literature bears emphasis. By and large, philosophers who take the trouble to argue that morality plays a constitutive role in well-being think it plays a *large* role—for Aristotle, e.g., “the central and controlling” role. So, for instance, an immoral person can’t be happy, which is the main point Plato tried to make with the *Republic*. If immorality only makes you a little less happy, or a little worse off, other things equal, so that the wicked can still be happy, then we’ve seemingly failed to vindicate morality in terms of self-interest. For example, an agent might face an important choice where she would be much better off choosing a life of ruthless chicanery, because the costs of immorality are outweighed by other prudential

[13] On ‘grue’, see Goodman (1983).

[14] In the case of happiness, see e.g. Haybron (2003, 2008).

[15] Some prominent examples from this period include Austin (1968), Hare (1963), Nozick (1989), and Smart (1973). For a review, see Uyl and Machan (1983). For criticism, see Haybron (2008). The situation arguably changed with Sumner (1996), which roughly marks the beginning of the contemporary philosophical debate.

goods like enjoyment and career success. This is not a prospect that should gladden the heart of Plato, Aristotle, or just about any other philosophical perfectionist of note.

At any rate, perfectionists standardly think the vicious cannot flourish—a far stronger claim than merely that vice is a little bad for you, other things equal.[16] While it is possible that some objective list theorists who take virtue to be just one among multiple constituents of well-being would be happy with the idea of evil flourishing, that view is not often, if ever, expressed in the literature. That morality is a better bet for happiness, *ceteris paribus*, is a weak claim of dubious significance.

We now turn to the experimental evidence, starting with an examination of Phillips et al.'s findings.[17]

3. Extant Findings

3.1 Experiment 1 by Phillips et al. (2017)

The first study of Phillips et al.'s widely cited paper reports a mixed design study. Participants were presented with one out of two scenarios (between-subjects) or both (within-subjects), in which the protagonists were described as enjoying high positive affect, little negative affect and a high overall satisfaction with their lives (henceforth the "scientific" conception of happiness (Diener, 2000; Diener, Scollon, & Lucas, 2004; Lucas, Diener, & Larsen, 2003; Zou, Schimmack, & Gere, 2013). The two agents differ in so far as one is moral and the other one is immoral. There were three scenarios (*Uncle, Nurse, Janitor*). In *Uncle*, a man either reads to his niece at night or else rapes her. In *Nurse*, a woman either helps sick children in the hospital or poisons them. In *Janitor*, a man either provides extra support to handicapped students or else steals belongings from students' lockers so as to resell them and buy alcohol. The two conditions of the *Nurse* scenario read:

> *Nurse*
> After going to nursing school for several years, Sarah got a job at the children's hospital and sees many different children each day. This is the job she has always wanted. Almost every single day Sarah feels good and generally experiences a lot of pleasant emotions. In fact, it is very rare that she would ever feel negative emotions like sadness or loneliness. When Sarah thinks about her life, she always comes to the same conclusion: she feels highly satisfied with the way she lives.
>
> [**Good:** The reason Sarah feels this way is that she helps the sick children by giving them vitamins that taste like gummy bears. / **Bad:** The reason Sarah feels this way is that she poisons the sick children by giving them vitamins that have pesticides inside of them.] Sarah doesn't really know how many children have [been helped by her / died because of her], but she likes to think about it when she falls asleep at night.

Following the scenario (or each scenario in the within-subjects design), participants had to respond to the question whether "[The protagonist] is happy." Responses were collected on a 7-point Likert scale, labelled "completely disagree" at 1, "in between" at 4, 'completely agree' at

[16] A prominent virtue ethicist, Rosalind Hursthouse, may seem to be an exception as she allows for the possibility that an immoral person can flourish, even if virtue is a person's best bet for flourishing (Hursthouse, 1999). But her account, like some others in the neo-Aristotelian line, defines virtue in terms of flourishing and not the other way around, in contrast to conventional readings of the Aristotelian tradition.

[17] Other recent studies of folk intuitions about happiness and well-being include, e.g., Brigard et al. (2010), Bronsteen et al. (forthcoming), Carlquist et al. (2017), Delle Fave et al. (2016), Hindriks & Douven (2018), Ip (2011), Joshanloo (2013, 2014), Mogilner et al. (2011), Oishi et al. (2013), Olivola et al. (2013), Pflug (2009), Sotgiu, (2016), Weijers (2014).

7. Participants also had to respond to true/false questions regarding the presence of positive affect, the absence of negative affect and a high overall life-satisfaction (the three features of the scientific conception of happiness, which are addressed in the first paragraph of the scenarios). Those who failed to answer any of the questions with "true" were excluded. This is because the authors aim to ensure "that any difference in responses was attributable to a moral difference but not a difference in whether the agent was viewed as having the scientifically agreed upon traits of happiness." (2017, p.169).

For the between-subjects experiment, a linear mixed effects model treating *moral character* and *academic expertise* regarding happiness as fixed factors, and *scenario* as a random factor revealed a significant main effect of moral character ($\chi^2(1)=28.71$, $p<.001$). Expertise and the interaction were nonsignificant (*ps>=.591*). In planned comparisons, the morally good agent (*M*=6.34, *SD*=1.09) was deemed significantly more happy than the morally bad agent (*M*=5.44, *SD*=1.62, *t*(212.50)=-5.2, *p*<.001, Cohen's *d* =.66). In the within-subjects design, there was again a significant main effect of moral character ($\chi^2(1)=48.65$, $p<.001$). The protagonists are rated happier when their character was specified as morally good (*M*=6.48, *SD*=.84) than when it was bad (*M*=5.63, *SD*=1.74), *t*(183)=7.44, *p*<.001, *d*=.55. The results are illustrated in Figure 1.

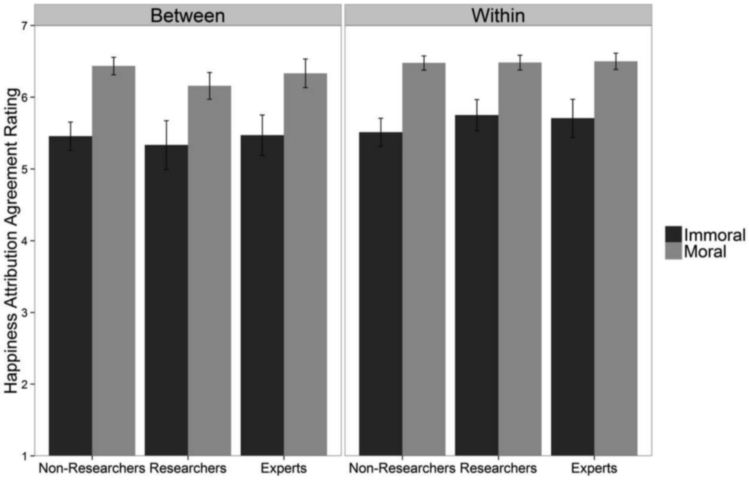

Figure 1: Participants' agreement that an agent is happy as a function of the participants' level of expertise and the moral value of the agent's life, which was manipulated both between-subjects (left) and within-subjects (right). Error bars indicate standard error of the mean. From Phillips et al. (2017, p. 170).

3.2 Discussion

For the moment, let's take the findings at face value: what do they mean? We ran one-sided t-tests that show that in both designs, the happiness of the evil characters was assessed as significantly above the midpoint of the scale (all *ps*<.001, see Appendix, Section 1.1) and substantially so (the means are about M=5.5 in both designs). One striking feature of the results is thus that the immoral agents are placed firmly in the "happy" category: the morally bad uncle may be deemed somewhat less happy than the good uncle, but he is considered happy nonetheless. In the eyes of the folk, then, Thrasymachus essentially stands vindicated, and the ancient eudaimonists and all those following them are wrong: happiness does not require virtue, and immoralists can perfectly well be happy. Not quite as happy, admittedly, as virtuous agents, but even that concession requires a *ceteris paribus* clause: the results suggest only that, *other things equal*, virtuous agents are somewhat happier than immoral ones. It is entirely possible that, in practice, other things are not equal: factors other than morality likely play a role in happiness as well, and it is entirely consistent with these results that in the real world evil people do sufficiently better on those counts to make them better off, on the whole, than good people. Maybe bad people tend to be richer, more admired, achieve more, and otherwise get more of what they want more than good people, so that the disadvantage of being immoral is outweighed by the benefits. We are not suggesting that these are plausible conjectures—probably all parties will agree that the world doesn't generally work like that. But philosophical perfectionists should be concerned if that is even a possibility.

In fact it is very unlikely that most philosophers who have posited a constitutive role for morality in happiness or well-being could derive a great deal of comfort from these data, at least from looking just at the averages: the effect of morality on happiness ascriptions is far too small to be squared with their theories. At best, the results may suggest ambivalence among the folk about the role of morality, so that they might be open to philosophical persuasion: on further reflection they would come to accept an Aristotelian or other perfectionist view. But this seems an improbable explanation of the data given that even philosophers who undertake such reflection for a living exhibit the same pattern as ordinary folk.

This does not mean that the results are without interest: there's a difference that needs explaining. One possibility is that morality plays a constitutive role in happiness that philosophers have overlooked, such as the "dual character" hypothesis suggested by Phillips et al. This is a provocative idea, but more needs to be said to explain the theoretical or practical value of understanding happiness in such a way, as we noted earlier.

Another possibility is that the averages mask important differences among the folk; once the data are disaggregated, perhaps it will turn out that a substantial portion of laypersons attribute a much larger role to morality in happiness, for instance deeming the immoral agents to be unhappy. We turn to that question in the following section.

4. Reexamining the evidence

The between-subjects data reported in Figure 1 is not surprising (see also the between-subjects results of Phillips et al. 2011 and Phillips et al. 2014). As discussed, moral valence has been found to have a strong impact on a wide range of judgments that seemingly would not (or should not) be sensitive to moral factors. By themselves, findings of this sort reveal little about the folk *concepts* at stake: they could be mere pragmatic phenomena (people wanting an outlet to voice their moral disapproval, see e.g. Adams & Steadman 2004a, 2004b, Steadman & Adams, 2007; Mizumoto, 2018) or they might testify to a blame-driven bias (Alicke, 2000, Alicke & Rose, 2010, Sauer & Bates, 2013). What *is* surprising, especially if one sympathizes with a pragmatic or a bias account, however, is the medium-sized morality effect on happiness attributions in the within-subjects design. If happiness were insensitive to moral factors, one would expect people *not* to judge the two agents differently in terms of happiness when the only relevant difference across scenarios, i.e. morality, stares them into the face.

But here too, one might be too quick to infer that *the* concept of happiness (or whatever "is happy" actually refers to) is sensitive to normative factors, and thus constitutes a dual character concept. The significant, yet only mid-sized difference in *mean* attribution of happiness across moral character might well be driven by a minority. We thus explored the proportion of participants who judged the immoral agent *differently* from the moral agent in the sample of Phillips et al. (2017), as only this percentage of subjects can genuinely be considered to use “is happy” in evaluative fashion. Across the three scenarios, the proportion was only about 33% (significantly below chance, binomial test, *p*<.001, two-tailed). Conversely, the proportion of participants who judged the happiness of the moral and the immoral agent *identically* across scenarios – thus manifesting *insensitivity* to moral concerns – was 67% (see Figure 2a), and significantly exceed the proportion who use “happy” in a morality-dependent fashion (binomial test, test proportion =.67, *p*<.001, two-tailed).

A closer look at the free text explanations for the responses collected by Phillips et al. suggests that many people very clearly reject the perfectionist conception of happiness. Here are a few examples:

> [Uncle] In either scenario, Garrett is happy about his choices. Happiness and reprehensible crime are not mutually exclusive.
> [Uncle] Different things can cause happiness. Just because a person is gross or a bad person doesn’t mean that they can’t feel happiness.
> [Nurse] If doing bad things makes you happy, then I guess you have to still be considered happy. Whether or not she is killing children or helping them. If it makes her happy then I guess it makes her happy. As weird as that sounds
> [Nurse] There is no difference to whether Sarah is happy, regardless of the morality of either scenario.
> [Janitor] Tom seemed to experience the same feelings of satisfaction and fulfillment regardless of the reasons.
> [Janitor] Tom is happy regardless of the moral consequences of the actions

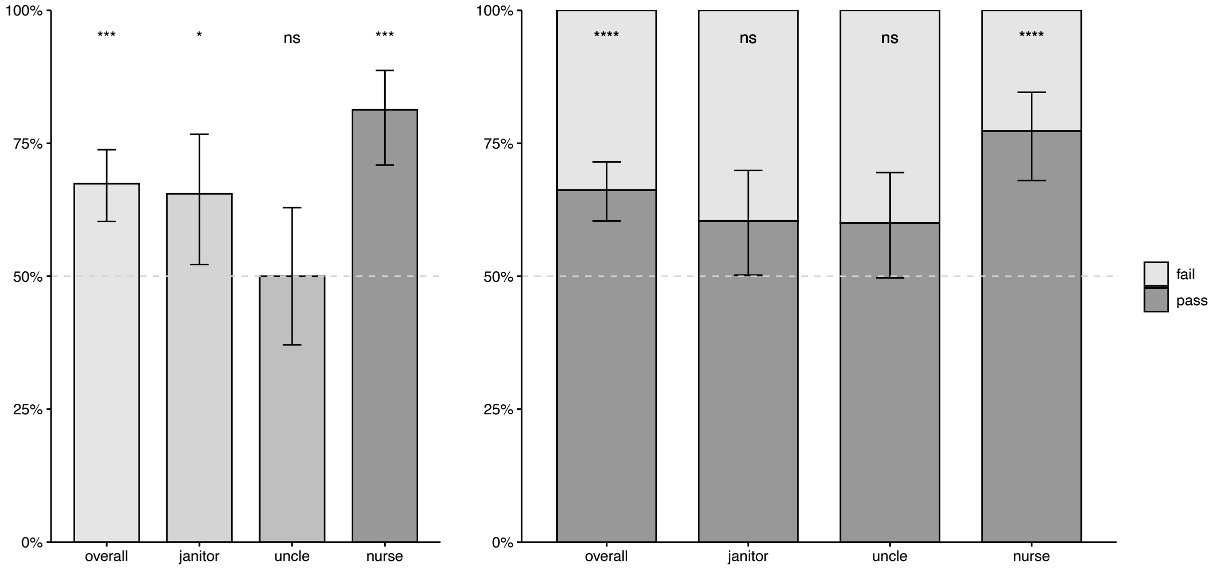

*Figure 2a: Proportion of participants who judge the two agents identically in terms of happiness. Error bars denote 95% confidence intervals. Figure 2b: Proportions of participants who failed / passed the comprehension check overall and across scenarios. Error bars denote 95% confidence intervals. Significance levels for the difference from chance (50%). *p<.05; **p<.01; ***p<.001; ****p<.001.*

At the level of individual scenarios, what stands out is that in the *Nurse* scenario about 80% of the people judged the two agents identically (manifesting insensitivity to moral factors), whereas in the *Janitor* and in particular the *Uncle* scenario, this proportion was considerably lower. In *Uncle*, for instance, only 50% judged the two agents as equally happy, which did not differ from chance (binomial test, *p*>.999). To explore why, it is instructive to look at the explanations provided by participants who judged the protagonist differently across moral conditions (responses by happiness experts are flagged with an asterisk):

> In both scenarios Garrett appears to be happy the problem lies when you try to determine what is happiness. The happiness that Garrett feel in the first scenario makes perfect sense to him but to us it makes us think that he has a **sick mind** so he doesn't know what happiness is. While the second scenario fits with most peoples view on happiness. So in the end both scenarios show happiness but only in the mind of Garrett.
>
> Garrett is a very **sick person** and can't really be happy. His life may appear happy, as in the first scenario, but after reading the second scenario, his life is not what it seems.
>
> The difference is that what he's doing in the second passage is complet[e]ly wrong. It's a physical release that he's **deluded** himself into thinking makes him happy.
>
> * Pressuring someone to have sex implies that there is **something missing in his life**, esp. if that person is a relative. Of course, he could just have some **disorder**, but I'm not a trained clinician. I'm mostly relying on my own moral judgment here.
>
> * The immoral version makes it feel like Garrett is **secretly unhappy (e.g. battling some inner demons)** and is **deceiving himself** when he thinks he is happy.

What many of these responses (laypeople and experts) suggest is that those who ascribe happiness to different degrees consider the immoral version of Garrett, the uncle, mentally unhealthy and/or delusional to an extent that he simply *cannot* enjoy true happiness. Interestingly, for this scenario the proportion of participants, who attributed high subjective well-being to both protagonists – and thus successfully completed the manipulation check – did not differ significantly from chance either (Figure 2b). This suggests that people had serious difficulties to take the scenario at face value. Even those who successfully completed the manipulation check, held views according to which "Garrett is a very sick person and can't really be happy". But this does not show that he is unhappy because he is *immoral*. Contrast the *Nurse* scenario, where more than 75% of the participants attribute high subjective well-being to a woman who murders innocent children (thus satisfying the manipulation check). This character seems to be judged as *merely immoral* (and not sick), and hence about 80% of the participants judge her as equally happy as her saintly counterpart.

The hypothesis just raised is confirmed by a closer look at the *Janitor* scenario. In principle, the bad janitor, Tom, is the least immoral agent of the three scenarios: Rather than raping and killing children, he steals their belongings from their lockers. What is surprising at first glance is that the proportion of participants who considered Tom as happy as his moral counterpart (about 60%) is much lower than in *Nurse*, and that the proportion of those who passed the manipulation check – and thus apparently succeed in taking the scenario at face value – is also much lower (also about 60%). However, here, too, a quick look at the explanations of the participants is instructive. In fact, participants take issue with the fact that Tom steals *in order to buy alcohol*. And yet again, this fits the hypothesis that they consider him sick to an extent that obstructs the possibility of genuine happiness:

> The difference is that in the first story, Tom steals to **feed his alcohol addiction**, which is hardly a positive way to go through life. He gets a high off of stealing, and then he gets **alcohol**. In the second story, Tom has reached a point of altruism.
>
> The fact that he is **an alcoholic probably means that he is unhappy** and this made it relevant to his happiness.
>
> *The difference was relevant because in the first passage it is demonstrated that he has a **drinking problem** that makes him commit an offence - stealing.
>
> *The first passage (about **alcohol use**) made me question whether or not Tom was truly happy. For example, many **drug addicts** say they are happy with their drug addicted lives, but once they become sober they reflect back on that time and say that they actually weren't happy, they were fooling themselves at the time.

In a nutshell, then, a closer look at the data gathered by Phillips et al. suggests two qualifications to their results: First, across the three scenarios, a strong majority (two in three participants) employ the expression "is happy" in a morality-insensitive fashion. Given that the significant effect in mean-attribution across moral character is driven by a minority, it is an exaggeration to suggest that *the* folk concept or conception of happiness is evaluative. Second, there is at least *some* reasons to suspect that among those who *did* judge the immoral and the moral agents differently in terms of happiness might do so on grounds that are not sufficiently controlled for: Upon closer inspection, especially the free text explanations for the *Uncle* and the *Janitor* scenarios, frequently make mention of mental illness, delusion or alcohol addiction. What this suggests is that two out of three vignettes harbor a potential confound: participants might think that there is something so fundamentally unwell with the mind of an incestuous rapists that people of this sort simply *cannot* be truly happy, as they lack the requisite mental states.

5. A New Experiment

A traditional difficulty in the philosophical debate over the connection between morality and happiness or well-being is that it is hard to devise cases that we can trust to elicit a clear signal from philosophers' intuitions. Generally, immoral people are widely thought, as a matter of contingent empirical fact, to suffer at some level from various disadvantages including mental illness or psychological distress—fear, doubt, guilt, loneliness, and so forth. The apparently cheerful drug lord may enjoy little peace of mind, always having to look out for betrayals, the police and so forth. A great portion of morality, probably the lion's share, concerns our dealings with family, friends and others in the local community. For intensely social creatures like ourselves it is bound to be hard to do well by any standard if we do not at least present as basically virtuous or decent agents—trustworthy, fair-minded, caring, and so forth. Bad people tend to have problems.

As a result, scenarios depicting seriously immoral individuals may typically generate expectations that the persons involved have those problems, making it hard to offer clear-cut counterexamples to the idea that happiness precludes a life of immorality. Given the case of a putatively happy drug lord, for instance, philosophers may and sometimes do simply argue that the stipulation isn't believable, and reject the claimed intuitions: we don't really have them, when we think it through.[18] Note that philosophical perfectionists who take this line are in no position to claim support from Phillips et al.'s studies: if they really distrust intuitions about bad people who

[18] E.g., Badhwar (2014).

are supposedly cheerful etc., then they should certainly distrust the intuitions in question here: by their own lights, the data don't show anything philosophically interesting.

There is no sure-fire way around these concerns, but they can at least be alleviated by turning to forms of immorality that are least likely to be associated with psychological distress and other problems. A natural candidate is the mistreatment of outsiders, as in tribal warfare or institutionalized slavery. Opinions may vary even here, but it plausibly tends to be easier for a psychologically normal human being to do terrible things to outgroup than ingroup individuals. (This is not an entirely trivial generalization. When a Kalahari Bushman was asked why killings in his tribe often involved family members, he replied "Why, who else would you want to kill?"[19] We don't think this affects the present discussion, but you can see his point.) When a form of immorality is supported by one's family, community, and culture, then most of the familiar reasons for expecting it to backfire may not apply. If your moral code and the people around you are united in endorsing or even rewarding the behavior, and may even sanction you for *not* engaging in it, then the personal costs of immorality are liable to be considerably lower, if there are costs at all. This is all the more plausible if one thinks, as many do, that human beings are well-suited if not wired for xenophobia and selective cruelty toward outsiders, as might be suggested by the astonishing rapidity and enthusiasm with which normally decent people have been known to embrace the most extreme forms of violence. So cases of this sort should provide a relatively good test of folk intuitions about the connection between happiness and morality: does immorality thereby compromise one's happiness? We expect that responses in this sort of case would yield a clearer signal than intuitions about within-family or within-community wrongdoing that violates local norms, such as molesting one's niece, poisoning one's patients, or stealing from students in one's workplace.

To this end we presented respondents with new scenarios. One pair of vignettes, for instance, is set in the antebellum South of the United States. The "immoral agent" condition involves a successful slaveholder, whereas the other condition describes a morally admirable but otherwise similar physician. How would judgments of happiness compare between the two characters?

5.1 Participants

We recruited 167 participants on Prolific Academic. In line with the preregistered criteria,[20] 3 participants who were not native English speakers or took less than twenty seconds to answer the main questions (including reading the prompt) were excluded. The final sample consisted of 164 participants (female: 52%; age M=39.2 years, SD=15 years, range: 35–81 years).

5.2 Methods and Materials

The experiment took a 2 scenario (between-subjects: Plantation, Warrior) x 2 moral character (within-subjects: good v. bad) mixed design. Participants were randomly assigned to one of the two scenarios. Both scenarios describe circumstances from the past in which deeply immoral behavior (owning and mistreating slaves, torturing outgroup members) was socially acceptable. The materials described two characters – one immoral, one moral – both of which enjoy a high

[19] Suzman (2017), p. 248.

[20] Preregistration: https://aspredicted.org/5TN_2HX. 21 out of a total of 188 participants could not take part in the survey because they failed one or more attention checks.

level of subjective well-being. Participants who received the *Plantation* scenario were presented with the following materials on a single screen:

> **Simon the plantation owner**
> It is the early 1800s, and Simon is a wealthy plantation owner in the Southern United States. He owns many slaves, who live in harsh conditions and are often treated cruelly. But this is typical of the region and Simon pays little attention to their fate.
>
> Simon enjoys a rich social life, a close family, and a leisurely lifestyle with interesting hobbies to engage his time. He is very satisfied with his life, takes great pleasure in everyday activities, and is generally cheerful and relaxed. He is untroubled, and sleeps well at night.
>
> **Bernard the doctor**
> It Is the early 1800s, and Bernar" is ' wealthy doctor in the Southern United States. He practices general medicine, and has cured many people in town and from other cities in the region who came to visit him.
>
> Bernard enjoys a rich social life, a close family, and a leisurely lifestyle with interesting hobbies to engage his time. He is very satisfied with his life, takes pleasure in many everyday activities, and is generally cheerful and relaxed. He is untroubled, and sleeps well at night.
>
> **To what extent do you agree or disagree with the following claims?** [1-completely disagree, 7-completely agree]
>
> [Morality] "Simon/Bernard is a morally good person." (manipulation check)
> [Happiness] "Simon/Bernard is happy."

The other scenario is situated in the 13th century and describes two men who belong to a tribe in the nomadic plains. It contrasts Joron, a warrior, who "is ruthless toward those outside his tribe, inspiring fear by subjecting men and women to torture after a successful raid" with Tunkur, a trader, who "treats outsiders with hospitality and respect, and when battle occurs, he shows mercy toward the defeated." Participants were also asked to what extent they agreed that Joron/Tunkur are morally good people and that they are happy (the complete materials are in the Appendix, Section 2.1).

Following the vignette-based assignment, participants were presented with a final task, in which we explicitly asked them whether, on their view, moral concerns should affect happiness attributions. The prompt read:

> We are interested in how you use the word 'happy'. In saying 'Mary is happy' some people mean that overall, Mary has a positive state of mind. For example: Mary generally feels good and experiences a lot of pleasant emotions. It is rare that she would feel negative emotions like sadness or loneliness. When Mary thinks about her life she is highly satisfied with it. This use of 'happy' is independent of moral factors such as whether Mary is a morally good person or not.
>
> Other people would only say that Mary is happy if, beyond having a positive state of mind, she is also a moral person. On this view, a person who, overall, has a positive state of mind but is immoral is not a happy person. This use of 'happy' depends both on a positive state of mind and on morality.
>
> What, to you, is the most appropriate way of using the word 'happy'?
> □ When I say Mary is happy, I only want to say that she has a positive state of mind.
> □ When I say Mary is happy, I want to say that she has a positive state of mind and is a morally good person.

The order of the responses was randomized. Following the explicit task, participants completed a brief demographic questionnaire.

5.3 Results

5.3.1 Vignette-based Task

The means for perceived moral character suggests that our manipulation was successful: Overall, the good characters are judged significantly more moral than the bad characters ($t(163)=23.73$, p <. 001, *Cohen's d* = 1.85, a very large effect, see Figure 3).

We estimated a linear mixed-effects model with the "lme4" package in R (Bates, Maechler, Bolker, & Walker, 2015), regressing the dependent variables on the moral characteristics of the people as fixed factor and scenario as a random factor to test the main effect of the moral characteristics. The analysis revealed a significant main effects on happiness ($\chi^2(1)=8.17$, $p=.004$). An ANOVA (type III) revealed that across scenarios moral character had a significant effect on happiness ($F(1,162)=19.88$, $p<.001$, $\eta_p^2=.025$, a small effect; resp. $F(1,163)=19.55$, $p<.001$, $\eta_p^2=.024$ without controlling for scenario, again a small effect). Planned contrasts revealed an overall small effect on happiness (*p<.001, d=.35*), see Figure 3. For the immoral characters, happiness ratings were significantly and substantially above the midpoint overall (mean rating close to 6 out of 7) and in each scenario (one-sample t-tests, all *ps*<.001). For details regarding the statistical Analyses, see Appendix, sections 2.2-2.4.

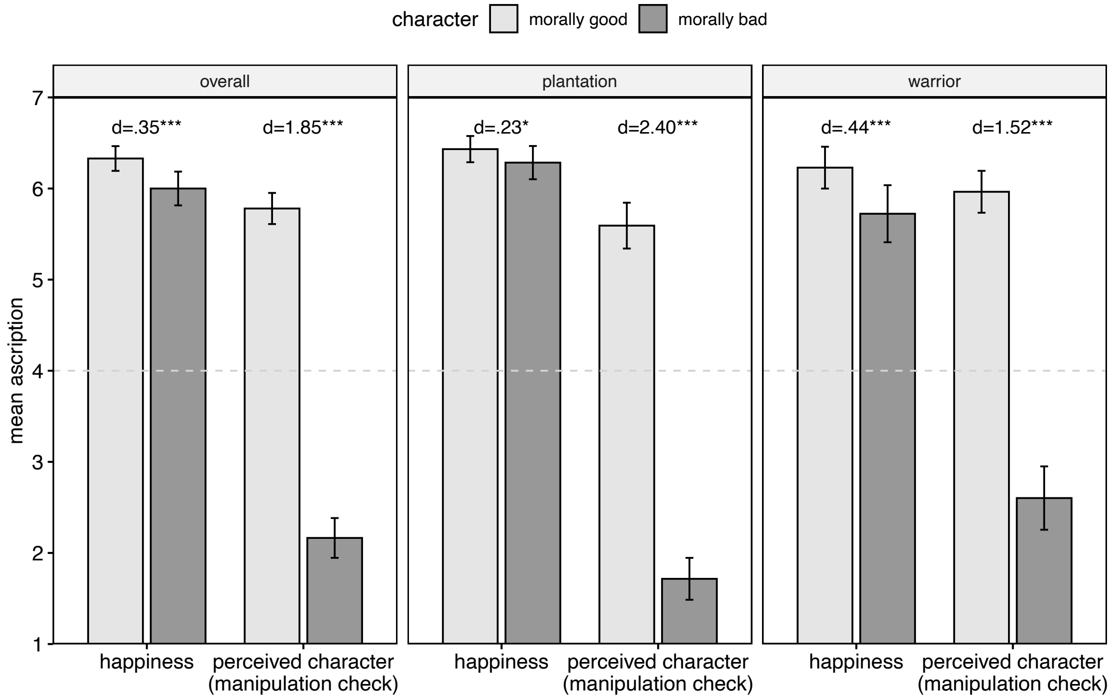

*Figure 3: Mean ascriptions for each DV across moral character overall and for each scenario. Error bars denote 95% confidence intervals. *p<.05; **p<.01; ***p<.001.*

Beyond the relatively small effect of moral character on perceived happiness, an analysis of the proportions of participants who judged the protagonists differently – and thus reveal that, in their view, moral character *should* make a difference – provides further insight. Overall, about four in five participants rated the happiness of the morally bad character as *identical* to the happiness of the morally good character (significantly above chance for each scenario and overall, all *ps*<.001, see Figure 4 and Appendix, section 2.5). This shows that the significant difference in mean happiness attribution across moral character is driven by a small minority of participants.

An explorative analysis revealed that the willingness to judge the two agents differently correlated negatively with religiosity, though the coefficient is small (*r*=-.21, *p*<.01). We found no significant correlations with age or gender (*ps*>.05, see Appendix, section 2.6).

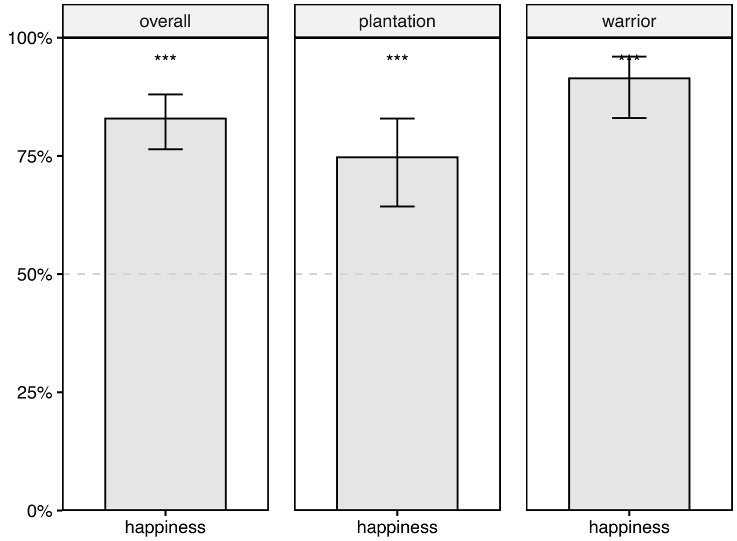

*Figure 4 Proportions of equal responses in the good and bad character condition overall and across scenarios. Error bars denote 95% confidence intervals. *p<.05; **p<.01; ***p<.001.*

5.3.2 Explicit Task

The results of the explicit task are near-identical to those of the implicit task. About four in five people responded that, for them, moral concerns do not play a role in the attribution of happiness, which significantly exceed chance overall and for each scenario (binomial tests, *p*<.001, see Appendix, section 2.7).

5.4 Discussion

Our results suggest that, once possible confounds due to mental illness, alcohol addiction and the like are ruled out, the mean effect of moral character on happiness attribution is small, and driven by a minority of participants (about 20%). Conversely, we found that about four in five participants implicitly and explicitly manifest a concept/conception of happiness insensitive to moral factors. Regarding the chief target of this paper, the folk concept or conception of happiness, the results suggest that morality plays little or no constitutive role in happiness as most commonly understood, consistently with the popular view that HAPPINESS is a purely psychological concept akin to the concepts of life satisfaction or tranquility. Also consistent with these results is the hypothesis that HAPPINESS is equivalent to WELL-BEING. In that case, our findings suggest that most of the folk see no necessary link between morality and self-interest. On either interpretation, our results appear to be of considerable philosophical interest.

Why did we obtain such different results from Phillips et al.? A partial answer is that we arguably did not: as we suggested earlier, the apparent effect of morality of happiness ascriptions in the earlier studies can largely be explained in nonmoral terms, with many participants insisting that the immoral individuals have some emotional disturbance that would credibly render them unhappy, or at least less happy, on a purely psychological, non-moralized conception of happiness. To some extent participants may simply have refused to accept the vignette at face value, but the focus in those vignettes on particular attitudes and emotions—feeling good, having lots of pleasant emotions, low negative emotions, and being highly satisfied with life—arguably leaves room for participants to fill in the blanks with familiar schemas involving unconscious, latent or backgrounded emotional distress, so that superficial “happiness” masks a fundamentally troubled mind. Some earlier work involving cases of that sort suggests that the folk may well be content to ascribe unhappiness on emotional grounds even to individuals experiencing predominantly positive feelings and judging their lives to be satisfactory, say if they are prone to break down in tears in the occasional quiet moment.[21] We think the most credible explanation of the Phillips et al. results is that, in the folk mind, well-adjusted and emotionally healthy people do not do things like poison their young patients. That issue appears to be considerably diminished in the present study both by its focus on locally “acceptable” immorality in a society organized around chattel slavery, and by the use of a familiar device for efficiently indicating internal harmony, being untroubled and sleeping well. (Recall that Plato starts the *Crito* with a remark about how soundly the doomed Socrates slept, immediately establishing that he enjoyed a high degree of emotional well-being.)

6. Conclusion

Phillips and colleagues pioneered a highly influential and important line of research that has brought much-needed attention to underexamined aspects of some of the central questions of philosophy and ordinary life: what does it mean to be happy? Philosophical work on this question cannot afford to be detached from the ways in which it enters into everyday thinking. Simply as a matter of respect, people’s ideas about the things that matter in life merit philosophical and empirical study, as does the question of what they mean when they talk about crucial goods like happiness. The lives in question are theirs, and one does not have to be a rank subjectivist to suspect that people are not entirely clueless about what matters in their lives. Here we build on the work of Phillips et al., even if we diverge from their conclusions.

The divergence is indeed substantial: we have argued that morality likely plays little or no role in the folk conception of happiness, at least for a majority of individuals, though our reanalysis of Phillips’ data makes room for some potential ambiguity of the expression “happy”. A likely explanation for the differing conclusions is that the original vignettes elicited pushback from participants, who might have found it implausible that the sorts of immorality posited could fail to exact a damaging emotional toll and either refused to accept the stipulation or filled gaps in the vignettes with forms of emotional distress or mental illness that had not clearly been ruled

[21] Haybron (2013).

out. This reading is supported by our findings in a study with new scenarios where some form of emotional distress was less likely to be assumed by participants. In these scenarios, the association between happiness and morality essentially vanished, and even judgments about well-being showed little impact from the blatant immorality of the character.

But even if one accepts Phillips et al.'s conclusions about the concept of happiness, we have argued that it is not clear what of philosophical significance follows: the posited effect of morality on happiness ascriptions is far weaker than most if not all philosophical theories claiming a constitutive link between morality and happiness require. Few welfare perfectionists would be content with the notion that Thrasymachus was essentially right: the grossly immoral can indeed be happy. On that supposition Plato, Aristotle and the rest of the tradition that followed lose: self-interest cannot be a sufficient basis for morality, because self-interest does not require it. Depending on how the facts of everyday life play out, self-interest may not even *counsel* morality.

A significant corollary of these reflections is that the philosophical significance of the posited link between happiness and morality is most readily apparent if happiness is also taken to be equivalent to well-being: when philosophers talk about the role of morality in "happiness," they almost invariably mean the role of morality in well-being. But that is a broader question requiring consideration of other lines of data besides the happiness-morality link, since well-being is usually taken to involve factors other than just morality and positive mental states. It would still be interesting if HAPPINESS instead took the form of a dual character concept with psychological and moral components—but how so, and whether philosophical work should adjust in light of such a finding, is not yet clear.

Some might find it disappointing that happiness, as the folk conceive it, bears no necessary connection to morality. Yet this result is entirely compatible with commonsense maintaining that, as things in human life actually are, happiness standardly counsels a high level of moral conduct. Happiness and morality can be very tightly coupled in practice even if entirely distinct in their natures; reliable associations do not require necessary connections. And as it happens, the qualitative reports from Phillips et al.'s studies suggest that people do tend to posit a significant positive connection between happiness and morality: it evidently takes some effort to come up with an example where they are willing to believe that serious immorality can fail to exact a psychological toll.

Bibliography

Adams, F., & Steadman, A. (2004a). Intentional action in ordinary language: Core concept or pragmatic understanding?. *Analysis*, *64*(2), 173-181.

Adams, F., & Steadman, A. (2004b). Intentional action and moral considerations: Still pragmatic. *Analysis*, *64*(3), 268-276.

Alexander, J. (2012). *Experimental philosophy: An introduction.* Cambridge: Polity.

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological bulletin*, *126*(4), 556.

Alicke, M., & Rose, D. (2010). Culpable control or moral concepts?. *Behavioral and brain sciences*, *33*(4), 330.

Annas, J. (1993). *The Morality of Happiness*. Oxford.

Austin, J. (1968). Pleasure and Happiness. *Philosophy*, *43*, 51–62.

Badhwar, N. K. (2014). *Well-Being: Happiness in a Worthwhile Life*. Oxford University Press.

Badhwar, N. K. (2015). Happiness. In G. Fletcher (Ed.), *The Routledge Handbook of Philosophy of Well-Being* (pp. 323–335). Routledge.

Baril, A. (2015). Virtue and Well-Being. In G. Fletcher (Ed.), The Routledge Handbook of Philosophy *of Well-Being* (pp. 242–258). Routledge.

Bengson, J. (2013). Experimental attacks on intuitions and answers. *Philosophy and Phenomenological Research, 86*(3), 495-532.

Besser, L. (2021). *The Philosophy of Happiness: An Interdisciplinary Introduction*. Routledge.

Brigard, F. D. (2010). If you like it, does it matter if it's real? *Philosophical Psychology*, *23*(1), 43–57. https://doi.org/10.1080/09515080903532290

Bronsteen, J., Leiter, B., Masur, J., & Tobia, K. (forthcoming). The Folk Theory of Well-Being. In S. Nichols & J. Knobe (Eds.), *Oxford Studies in Experimental Philosophy, Volume 5* (p. 18).

Buckwalter, W. (2022) A Guide to Thought Experiments in Epistemology (ms).

Carlquist, E., Ulleberg, P., Fave, A. D., Nafstad, H. E., & Blakar, R. M. (2017). Everyday Understandings of Happiness, Good Life, and Satisfaction: Three Different Facets of Well-being. *Applied Research in Quality of Life*, *12*(2), 1–25. https://doi.org/10.1007/s11482-016-9472-9

Darwall, S. (2002). *Welfare and Rational Care*. Princeton University Press.

Delle Fave, A., Brdar, I., Wissing, M. P., Araujo, U., Castro Solano, A., Freire, T., Hernández-Pozo, M. D. R., Jose, P., Martos, T., Nafstad, H. E., Nakamura, J., Singh, K., & Soosai-Nathan, L. (2016). Lay Definitions of Happiness across Nations: The Primacy of Inner Harmony and Relational Connectedness. *Frontiers in Psychology*, *7*(16), 410–424. https://doi.org/10.3389/fpsyg.2016.00030

Díaz, R., & Reuter, K. (2021). Feeling the right way: Normative influences on people's use of emotion concepts. *Mind & Language*, *36*(3), 451–470. https://doi.org/10.1111/mila.12279

Diener, E. (2000). Subjective well-being. The science of happiness and a proposal for a national index. *American Psychologist, 55,* 34–43.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49,* 71–75.

Diener, E., Scollon, C. N., & Lucas, R. E. (2004). The evolving concept of subjective well-being: The multifaceted nature of happiness. In P. T. Costa & I. C. Siegler (Eds.), *Advances in cell aging and gerontology* (Vol. 15, pp. 187–220). Amsterdam, the Netherlands: Elsevier.

Foot, P. (2001). *Natural Goodness*. Oxford University Press.

Fuhrer, J., & Cova, F. (2022). What makes a life meaningful? Folk intuitions about the content and shape of meaningful lives. *Philosophical Psychology*, 1-33.

Hare, R. M. (1963). *Freedom and Reason*. Oxford.

Haybron, D. M. (2003). What do we want from a theory of happiness? *Metaphilosophy*, *34*(3), 305–329.

———. (2008). *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*. Oxford University Press.

———. (2013). *Happiness: A Very Short Introduction*. Oxford University Press.

Heathwood, C. (2021). *Happiness and Well-Being*. Cambridge University Press.

Hindriks, F., & Douven, I. (2018). Nozick's experience machine: An empirical study. *Philosophical Psychology*, *31*(2), 1–21. https://doi.org/10.1080/09515089.2017.1406600

Horvath, J. (2013). How (not) to react to experimental philosophy. In *Experimental Philosophy and its Critics* (pp. 177-210). Routledge.

Horvath, J., & Wiegmann, A. (2016). Intuitive expertise and intuitions about knowledge. *Philosophical Studies*, *173*, 2701-2726.

Horvath, J., & Wiegmann, A. (2022). Intuitive expertise in moral judgments. *Australasian Journal of Philosophy*, *100*(2), 342-359.

Hursthouse, R. (1999). *On Virtue Ethics*. Oxford University Press.

Ip, P. K. (2011). Concepts of Chinese folk happiness. *Social Indicators Research*, *104*(3), 459–474. https://doi.org/10.1007/s11205-010-9756-7

Joshanloo, M. (2013). A Comparison of Western and Islamic Conceptions of Happiness. *Journal of Happiness Studies*, *14*(6), 1857–1874. https://doi.org/10.1007/s10902-012-9406-7

Joshanloo, M. (2014). Eastern Conceptualizations of Happiness: Fundamental Differences with Western Views. *Journal of Happiness Studies*, *15*(2), 475–493. https://doi.org/10.1007/s10902-013-9431-1

Kauppinen, A. (2007). The rise and fall of experimental philosophy. *Philosophical explorations, 10*(2), 95-118.

Keltner, D. (2009). *Born to Be Good: The Science of a Meaningful Life*. W. W. Norton & Company.

Kesebir, P., & Diener, E. (2014). A Virtuous Cycle: The Relationship Between Happiness and Virtue. In N. E. Snow & F. V. Trivigno (Eds.), *The Philosophy and Psychology of Character and Happiness* (pp. 287–306). Routledge.

Kneer, M., & Haybron, D. (2019). *Happiness and Well-Being: Is It All in Your Head? Evidence from the Folk*. https://doi.org/10.13140/RG.2.2.12958.69448

Kneer, M., Colaço, D., Alexander, J. & Machery, E. (2021). On second thought: Reflections on the reflection defense. *Oxford Studies in Experimental Philosophy Volume 4,* 257-296.

Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and brain sciences*, *33*(4), 315-329.

Knobe, J., Prasada, S., & Newman, G. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition*, *127*, 242–257.

Liao, S. M. (2008). A defense of intuitions. *Philosophical Studies, 140*(2), 247-262. Ludwig, K. (2007). The epistemology of thought experiments: First person versus third person approaches. *Midwest Studies in Philosophy, 31*(1), 128-159.

Lucas, R. E., Diener, E., & Larsen, R. J. (2003). Measuring positive emotions. In S. J. Lopez & C. R. Snyder (Eds.), *Positive psychological assessment: A handbook of models and measures* (pp. 201–218). Washington, DC: American Psychological Association.

Machery, E. (2011). Thought experiments and philosophical knowledge. *Metaphilosophy*, *42*(3), 191-214.

Machery, E. (2017). *Philosophy within its proper bounds.* London: Oxford University Press.

Mackie, J. (1990). *Ethics: Inventing right and wrong*. Penguin UK.

Mizumoto, M. (2018). A simple linguistic approach to the Knobe effect, or the Knobe effect without any vignette. *Philosophical Studies*, *175*(7), 1613-1630.

Mogilner, C., Kamvar, S. D., & Aaker, J. (2011). The Shifting Meaning of Happiness. *Social Psychological and Personality Science*, *2*(4), 395–402. https://doi.org/10.1177/1948550610393987

Nado, J. (2015). Intuition, philosophical theorizing, and the threat of skepticism. *Experimental Philosophy, Rationalism, and Naturalism: Rethinking Philosophical Method*, 204.

Nado, J. (2016). The intuition deniers. *Philosophical Studies*, *173*(3), 781-800.

Newman, G. E., De Freitas, J., & Knobe, J. (2015). Beliefs about the true self explain asymmetries based on moral judgment. *Cognitive Science, 39,* 96–125

Nichols, S., & Ulatowski, J. (2007). Intuitions and individual differences: The Knobe effect revisited. *Mind & Language*, *22*(4), 346-365.

Nozick, R. (1989). *The Examined Life*. Simon and Schuster.

Oishi, S., Graham, J., Kesebir, S., & Galinha, I. C. (2013). Concepts of Happiness Across Time and Cultures. *Personality and Social Psychology Bulletin*, *39*(5), 559–577. https://doi.org/10.1177/0146167213480042

Olivola, C. Y., Machery, E., Cheon, H., Kurniawan, I. T., Mauro, C., Struchiner, N., & Susianto, H. (2013). *Reality Does Not Bite* (pp. 1–13).

Pflug, J. (2009). Folk Theories of Happiness: A Cross-Cultural Comparison of Conceptions of Happiness in Germany and South Africa. *Social Indicators Research*, *92*(3), 551–563. https://doi.org/10.1007/s11205-008-9306-8

Phillips, J., De Freitas, J., Mott, C., Gruber, J., & Knobe, J. (2017). True happiness: The role of morality in the folk concept of happiness. *Journal of Experimental Psychology: General*, *146*(2), 165.

Phillips, J., Misenheimer, L., & Knobe, J. (2011). The ordinary concept of happiness (and others like it). *Emotion Review*, *3*(3), 320-322.

Phillips, J., Nyholm, S., & Liao, S. Y. (2014). The good in happiness. In *Oxford studies in experimental philosophy: Volume 1* (pp. 253-293). Oxford University Press.

Prinzing, M., & Fredrickson, B. (2022). *No Peace for the Wicked? Immorality is (Usually) Thought to Disrupt Intrapersonal Harmony* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/ug8tk

Prinzing, M., Knobe, J., & Earp, B. D. (2022). *Why Moral Judgments Affect Happiness Attributions: Testing the Fittingness and True Self Hypotheses* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/5dkp3

Raibley, J. R. (2012). Happiness is not Well-Being. *Journal of Happiness Studies*, *13*(6), 1105–1129. https://doi.org/10.1007/s10902-011-9309-z

Reuter, K. (2019). Dual character concepts. *Philosophy Compass*, *14*(1), e12557.

Reuter, K., Messerli, M., & Barlassina, L. (2022). Not more than a feeling: An experimental investigation into the folk concept of happiness. *Thought: A Journal of Philosophy*.

Ricard, M. (2015). *Altruism: The power of compassion to change yourself and the world*. Little, Brown and Co.

Rossi, M., & Tappolet, C. (2016). Virtue, Happiness, and Well-Being. *The Monist*, *99*(2), 112–127.

Sauer, H., & Bates, T. (2013). Chairmen, cocaine, and car crashes: The Knobe effect as an attribution error. *The Journal of ethics*, *17*(4), 305-330.

Smart, J. J. C. (1973). An Outline of a System of Utilitarian Ethics. In J. J. C. Smart & B. Williams (Eds.), *Utilitarianism: For and Against* (pp. 3–74). Cambridge University Press.

Sotgiu, I. (2016). Conceptions of Happiness and Unhappiness among Italian Psychology Undergraduates. *PLoS ONE*, *11*(12), e0167745-16. https://doi.org/10.1371/journal.pone.0167745

Steadman, A., & Adams, F. (2007). Folk concepts, surveys and intentional action.

Sumner, L. W. (1996). *Welfare, Happiness, and Ethics*. Oxford University Press.

Suzman, J. (2017). *Affluence Without Abundance*. Bloomsbury Publishing.

Sytsma, J. (2019). The character of causation: Investigating the impact of character, knowledge, and desire on causal attributions. (ms).

Tiberius, V. (2008). *The reflective life: Living wisely with our limits*. Oxford.

Tiberius, V. (2015). *Moral Psychology: A Contemporary Introduction*. Routledge.

Toner, C H. "Aristotelian Well-Being: A Response to LW Sumner's Critique." *Utilitas* 18, no. 3 (January 1, 2006): 218.

Uyl, D. D., & Machan, T. R. (1983). Recent Work on the Concept of Happiness. *American Philosophical Quarterly*, *20*(2), 115–134.

Weijers, D. (2014). Nozick's experience machine is dead, long live the experience machine! *Philosophical Psychology*, *27*(4), 513–535. https://doi.org/10.1080/09515089.2012.757889

Weinberg, J. M., Alexander, J., Gonnerman, C., & Reuter, S. (2012). Restrictionism and reflection: Challenge deflected, or simply redirected? *The Monist, 95*(2), 200-222.

Weinberg, J. M., Gonnerman, C., Buckner, C., & Alexander, J. (2010). Are philosophers expert intuiters?. *Philosophical Psychology*, *23*(3), 331-355.

Williamson, T. (2007). *The philosophy of philosophy*. Oxford: Blackwell.

Yang, F., Knobe, J., & Dunham, Y. (2021). Happiness is from the soul: The nature and origins of our happiness concept. *Journal of Experimental Psychology: General*, *150*, 276–288. https://doi.org/10.1037/xge0000790

Zou, C., Schimmack, U., & Gere, J. (2013). The validity of well-being measures: A multiple-indicator-multiple-rater model. *Psychological As- sessment, 25,* 1247–1254.