Jonathan Phillips

Sven Nyholm

Shen-yi Liao

The Good in Happiness

for *Oxford Studies in Experimental Philosophy,* Volume I

Abstract

There has been a long history of arguments over whether happiness is anything more than a particular set of psychological states. On one side, some philosophers have argued that there is not, endorsing a *descriptive* view of happiness. Affective scientists have also embraced this view and are reaching a near consensus on a definition of happiness as some combination of affect and life-satisfaction. On the other side, some philosophers have maintained an *evaluative* view of happiness, on which being happy involves living a life that is normatively good. Within the context of this debate we consider how people ordinarily understand *happiness*, and provide evidence that the ordinary understanding of happiness reflects aspects of both evaluative and descriptive views. Similar to evaluative views, normative judgments have a substantive role in the ordinary understanding of happiness. Yet, similar to descriptive views, the ordinary understanding is focused on the person's psychological states and not the overall life they actually lived. Combining these two aspects, we argue that the ordinary understanding of happiness suggests a novel view on which happiness consists in experiencing positive psychological states when one ought to. This view, if right, has implications for both philosophical and psychological research on happiness.

The study of happiness has recently benefited from a wealth of scientific research. One important development in this research has been an emerging consensus on the definition of happiness employed by scientific researchers across a number of different fields. This scientific definition of happiness concerns three core psychological components: high positive affect, low negative affect, and high life satisfaction (for a review, see Diener, Scollon, & Lucas 2003; Lucas, Diener & Lawson 2003). On this definition of happiness, whether or not a person is happy depends solely on that person's psychological states.

In contrast to this definition of happiness, a tradition in philosophy has long held that whether or not a person is happy actually depends in large part on whether the person is living a life that is *good*. An early form of this view can be found in Aristotle (*NE* 1097b22 – 1098a20), and is supported by several contemporary philosophers as well. What both ancient and contemporary versions of these theories have in common is that they hold that happiness is more than just a matter of which psychological states a person experiences. If a person does not have a life that is good, she will not truly be happy, no matter what she may subjectively experience.

Given the stark contrast between these two conceptions of happiness, one question that may naturally arise is how people *ordinarily* understand happiness when they lack the theoretical commitments that exist on both sides of this debate. Is the ordinary understanding concerned with only the psychological states a person experiences in accordance with the scientific definition, or does it place an emphasis on the value of the life the person is living?[1] This is the question we attempt to answer.

We begin by briefly reviewing some key ideas representative of the philosophical tradition that understands happiness as essentially normative or evaluative. Then we consider the anti-evaluative counter-reaction to this tradition amongst both philosophers and psychologists who propose to understand happiness in a purely descriptive, psychological way. Turning from these philosophical and psychological accounts, we present six studies

---

[1] By 'ordinary' we mean to pick out those uses of the word 'happiness' that are specifically not used in a technical way. One could stipulate a technical sense of 'happiness' to refer to *just* the psychological states or *just* the value of a person's life (as psychologists and philosophers have done), but those uses are not particularly interesting for our purposes. One could, after all, stipulatively use the term 'happiness' to refer to almost anything. What we are interested in is the ordinary, non-stipulative use of the word 'happiness.' It is, of course, possible that such concepts admit of variation between individuals, cultures and languages. For our purposes, though, we will set this variance aside, and consider the English word 'happiness' as a case study. We take up the question of whether there may be more than one single understanding of 'happiness' in Section V.

that investigate how people ordinarily make judgments of happiness. Five of these studies are reported for the first time, and we summarize one additional study (Study 3) that has been previously published (Phillips, Misenheimer & Knobe 2011). Together, these studies suggest that the ordinary understanding of happiness actually combines both evaluative and descriptive components while differing in important respects from both the evaluative and descriptive theories of happiness that have been previously proposed. Very roughly put, our analysis is that happiness, as it is ordinarily understood, consists in experiencing positive mental states when one ought to experience them. After proposing a general theoretical account of the ordinary understanding of happiness, we end by situating this research within the contemporary philosophical and psychological literature.

I.   *The philosophy of happiness*

Philosophy bears a long history of arguments over whether there is anything more to happiness than enjoyable psychological states. On one side, some philosophers (much like contemporary psychologists) have held that all there is to happiness is the possession of certain psychological states, e.g., pleasure or satisfaction. This position can be thought of as *descriptive*; it suggests that questions of whether a person is happy can be settled by simply considering an accurate description of the psychological states a person experiences. If the person has the requisite psychological states, then that person is happy. On the other side of this debate, philosophers have also maintained that being happy involves living a life that is normatively good. This second position can be thought of as *evaluative*; it suggests that questions about whether a person is happy cannot be settled with just a description of the persons' experiences or life – the important issue is whether that life is genuinely good. If the person's life is not good, then that person will not truly be happy, no matter how pleasant her psychological states may be. A number of different versions of both of these families of views have been proposed, and it will be helpful to get a rough grasp on how they are related to and differ from one another.

At the outset, theories of happiness should be distinguished from theories of well-being. Theories of well-being are concerned with what is ultimately best for a person in a non-instrumental way. Thus, hedonist theories argue that what is non-instrumentally best for a person is some form of hedonic pleasure, and objective list theories argue that there is some objective list of what is non-instrumentally best for a person, and argue over what

should be included on this list, and so on. As such, theories of well-being need not be directly concerned with what happiness is. Accordingly, we focus the remainder of our discussion specifically on theories of happiness, as these theories *do* directly concern what happiness is.[2]

Evaluative happiness

An early example of a view on which happiness requires more than the possession of particular psychological states can be found in Plato's *Gorgias* where Socrates mocks Callicles for believing that the mere satisfaction of desires is happiness. Offering a pointed counterexample, Socrates asks Callicles if he thinks that one could become happy just by continually scratching an itch (494c – e). Similarly, Aristotle argues that only the most vulgar people suppose happiness to be pleasure – the type of people who would choose lives better suited for grazing cattle (*NE* 1095b14 – 1095b21) and goes on to argue that happiness consists in living a virtuous life (*NE* 1097b22 – 1098a20).[3] While Plato and Aristotle mark out positions which emphasize virtue, scholars have disputed whether these ancient philosophers were actually concerned with happiness, suggesting the term *eudaimonia* would often be better translated as 'flourishing' or 'well-being' (Cooper 1987).

Yet we do not need to look to the ancient debate to find this position; it can also be found amongst contemporary philosophers who have been quite clear that they *are* concerned with happiness. To take one example, Philippa Foot has argued that 'when we talk about a happiness that is supposed to be a human good, we cannot intend pleasure or contentment alone.' Instead, she continues, 'there is a kind of happiness which only goodness can achieve' (2001). In support of this view, Foot asks somewhat incredulously whether a Nazi war criminal who escaped to Brazil after overseeing a death camp could really be happy (2001, 90-96). Her answer, while nuanced, is that it should be clear that he is not truly happy.

---

[2] It is worth noting that the one way in which theories of well-being may directly concern happiness is if the theory of well-being argues that what is non-instrumentally best for a person is to be happy. This position is clearly compatible with any theory of happiness.

[3] There is a question of whether Aristotle's continued discussion of happiness in Book X of Nicomachean Ethics suggests that the happy life is not necessarily a *morally* virtuous life (Cooper 1987), but the discussion of that question is outside the purview of this paper. For our purposes we will identify a broadly 'Aristotelian' conception of happiness with one that involves living a morally virtuous life.

In another example meant to illustrate the evaluative nature of happiness, Richard Kraut imagines a man who wants most to be loved and admired by his friends (1979). Upon learning about his wishes, his friends collude to trick him into believing that they love and admire him, even though they do not. The man comes to believe that his greatest wishes have been fulfilled and experiences great pleasure and satisfied with his life. Considering whether this man is happy, Kraut's response is as unambiguous as Foot's.[4] They both argue that happiness must be more than the mere sum of a person's psychological states.

Regardless of the merits of any one of these positions, what holds this diverse group of theories together is the criteria they use to determine whether or not a person is happy: they require an answer to whether that person's life is good.

Descriptive happiness

In contrast, other philosophers have pursued the study of happiness as a purely descriptive project. In this tradition, normative questions about the value of a person's life are simply irrelevant to whether or not a person is happy. Amongst the ancients, we can also find views not very dissimilar to this. Consider the Cyrenaics, who held that happiness among other goods could be found only in pleasure itself, even when it was gained at the expense of others' welfare. In an amusing anecdote meant to illustrate this position, Aristippus and Plato were said to be at the court of Dionysius, who suggested that they amuse themselves by putting on women's clothing. In the heedless pursuit of pleasure, Aristippus readily humiliated himself, while Plato was predictably much less fun (Giannantoni frs. 25-43).[5]

In the more contemporary literature on happiness, there are three main families of descriptive theories: *hedonist* theories, *life-satisfaction* theories, and *hybrid* theories.[6] Hedonist views of happiness suggest that one is happy if on balance, one experiences more pleasure

---

[4] One important difference between these two theories is that Kraut emphasizes the importance of the man's own values, whereas Foot does not. However both views are clearly evaluative in the relevant sense.

[5] Perhaps more cynically, Aristippus was also rumored to have spurned his own child as he was of no use to him in attaining pleasure (Diogenes II 81). Aristippus was only loosely associated with the Cyrenaics and he most likely did not contribute to their later written works as he predated the school by a number of years. Regardless, the example was taken to be illustrative of the Cyrenaics position, even amongst the ancients (Annas 2011), though the truth of the story has also been disputed (Annas 1993).

[6] A fourth family may be emotional-state views (Haybron 2008; 2010), though this family of views is relatively new. Similar to hedonist, life-satisfaction and hybrid theories, emotional-state theories take a purely descriptive approach to happiness.

than pain (e.g., Feldman 2010).[7] Life-satisfaction views, in contrast, focus on the attitude one has about one's life as a whole, and equate happiness with a reflective endorsement of the life that one is leading (e.g., Kekes 1982; Suikkanen 2011). And hybrid theories attempt to combine elements from both of these views (e.g., Sumner 1996; 2000).

For our purposes, what is most relevant about these theories is not the differences between them, but what they all have in common. All of these theories identify happiness with some psychological state or set of psychological states. They are all descriptive theories of happiness in the relevant sense because, according to them, we can determine whether or not a person is happy based only on an accurate description of the agent's psychological states.

II. *The psychology of happiness*

In contrast to the debate within philosophy, the psychological research on happiness has for the most part solely pursued the descriptive side of the project,[8] seeking both to define happiness as a set of psychological states and create methods for measuring whether or not people possess those psychological states (for a review, see Seligman & Csikszentmihalyi 2000). This understanding of happiness combines aspects of hedonic theories which emphasize pleasure and pain, and life-satisfaction theories which emphasize global attitudes about one's life (Diener, Scollon, & Lucas 2003). There is quite a widespread consensus that happiness can be best captured by a purely descriptive theory, which is often made explicit (e.g., Gilbert 2006, 36-7). As Gilbert and others make clear, to know whether or not a person is happy, one need do no more than come up with unbiased measures of that person's psychological states, whether he be a mass murder or Mahatma Ghandi.[9]

In the next section we turn to consider the ordinary understanding of happiness. Before doing so though, it may be helpful to briefly step back and consider how these various positions relate to one another. Early on, we distinguished theories of happiness from

---

[7] Fred Feldman's 'Attitudinal Hedonism about Happiness' is actually somewhat nuanced. On his theory, 'to be happy at a time is to have a positive net balance of intrinsic attitudinal pleasure over intrinsic occurrent displeasure at that time' or more simply, 'to be happy at a moment is to be taking more pleasure than displeasure in things at that moment' (2010). Thus, whether one is happy is a descriptive psychological question which depends on whether one takes more pleasure than displeasure in things at a given moment.
[8] While a number of psychologists have been interested in well-being as opposed to happiness, e.g., Ryan and Deci's (2000; 2012) self-determination theory and Ryff and Singer's multidimensional account (1998), few if any have offered genuinely evaluative theories of happiness.
[9] For one exception which proves the rule, see Kashdan, Biswas-Diener & King (2008).

theories of well-being, which are not concerned with the happiness of the agent, but rather with what is non-instrumentally best for the agent. Then, focusing on theories of happiness, we considered both evaluative and descriptive theories. Evaluative theories, on the one hand, hold that whether or not an agent is happy depends on whether or not the agent is living a normatively good life. Descriptive theories, on the other, hold that whether or not an agent is happy depends *only* on the particulars of the psychological states the agent experiences – though different descriptive theories choose to emphasize different psychological states.[10]

Importantly, these two views of happiness differ along two distinct dimensions. The first is that they differ over which aspect of an agent's life they focus on when determining whether or not the agent is happy. Evaluative theories of happiness tend to emphasize the agent's actual life rather than solely the agent's psychological states when determining whether the agent is happy. Descriptive theories, in comparison, tend to focus solely on the agent's psychological states. Secondly, these two views of happiness differ in the role they give normative judgments. Notice that some form of normative judgment may actually play a part in both evaluative *and* descriptive theories. Evaluative theories of happiness take the relevant normative judgments to be about how the agent's life is *actually going*, independent of anything the agent might think or believe. By contrast, descriptive theories of happiness hold that this sort of normative judgment is completely irrelevant to whether or not the agent is happy, as happiness can be determined by simply considering which psychological states the agent experiences. However, a certain sort of normative judgment may still play a role in descriptive theories in that some of the agent's mental states may *involve* normative judgments. For example, some purely descriptive theories of happiness hold that an agent must be satisfied with her life to be happy, and this psychological state of life-satisfaction presumably involves a normative judgment that the agent makes about how her life is going. For our purposes though, it is only evaluative theories of happiness that provide a genuine role for normative judgment, because on these theories, one simply cannot determine how happy a person without first determining whether or not they actually lead a good life. We return to this distinction between these two roles for normative judgments at the end of the

---

[10] Almost all further distinctions made by psychologists, e.g., between Kahneman's 'experiencing self' and 'remembering self' (Kahneman 2011) fall squarely within this framework as specific ways of working out descriptive theories of happiness.

paper, and argue that it is key to fully grasping how the ordinary understanding of happiness fits within the long-standing debate over happiness.

III. Ordinary happiness

The amount of debate over the true nature of happiness is perhaps only rivaled by the even more striking consensus about what people ordinarily mean by the English term 'happiness.' One finds the suggestion that 'happiness' refers to a descriptive psychological condition not just among psychologists, but among philosophers as well. Fred Feldman effectively illustrates the widely-held position:

> I think 'happy' is used in the evaluative sense only by philosophers and others whose good linguistic instincts have been perverted by them… As a result of this, I am inclined to treat such uses of 'happy' as semi-stipulative, semi-technical uses. It is not clear to me that ordinary speakers of English, completely innocent of philosophy, ever use 'happy' to mean the same things as 'having sufficiently high welfare.' (2010, 136)[11]

Indeed, there is no shortage of researchers who have made similar points (Cooper 1987; Crisp 2008; Feldman 2010; Haybron 2008; 2011; Sidgwick 1907/1966).[12] Yet despite the widespread agreement about ordinary usage, there has been surprisingly little empirical research on the matter. As of writing, we are currently aware of only one informal survey outside of the authors' own research (Haybron 2008).[13] We propose to fill this oft-overlooked lacuna with a series of systematic studies that actually investigate how people ordinarily understand happiness.

Moral concepts

---

[11] Feldman continues: 'A final comment: the OED reminds us that we sometimes speak of "happy coincidences," where a happy coincidence is just a lucky, or fortunate coincidence. The OED also says that 'happy' can be used to mean 'drunk.' I am inclined to think that these claims may be right. Perhaps the word does have these two additional senses when used in ordinary English. But again, I am not aware of a decisive test that would put either of these claims beyond question.'

[12] There are also a few rare examples of arguments against these claims, see, e.g., Dybikowski (1981) and Kraut (1979).

[13] There are, however, now a number of research programs on the topic that have begun, e.g., Vonasch & Baumeister (2012).

Outside of philosophical and psychological arguments over happiness, there is some independent reason to think the ordinary concept of happiness might involve normative evaluations. A growing body of research has demonstrated that moral judgments systematically influence how people ordinarily understand a number of concepts that were initially thought to be purely descriptive.

To illustrate, consider how it is that people ordinarily go about determining whether or not someone values something. One view would be that people have a purely descriptive understanding of valuing in which they simply consider a description of the agent's psychological states and decide whether any of those states count as the requisite pro-attitude (some positively valenced psychological state toward something under a particular description, e.g., thinking it good). An alternative view would be that people have an evaluative way of determining whether someone values something. On this sort of view, judgments of whether or not someone values something depend at least in part on whether the thing *itself* is genuinely good. Attempting to decide between these two possibilities, Knobe and Roedder (2009) asked participants to read a vignette about a person who held a particular psychological attitude toward the action of refraining from premarital sex and asked them whether the person valued that action. Surprisingly, participants' answer to this question depended on whether or not they happened to believe that refraining from premarital sex was a good thing.[14]

Surprising or otherwise, the evidence in favor of evaluatively-laden concepts does not stop there. To take just a few other examples, moral judgments seem to play a systematic role in the folk concepts of *causation* (Cushman, Knobe & Sinnott-Armstrong 2008; Hitchcock & Knobe 2009; Roxborough and Cumby 2009), *deciding* (Petit & Knobe 2009), *freedom* (Phillips & Knobe 2009; Young & Phillips 2011), *intentional action*, (Knobe 2003), *knowledge* (Beebe & Buckwalter 2010), and *weakness of will* (May & Holton 2011), and new discoveries continue to be made.

---

[14] In this study, participants were asked to read a vignette about a woman named Susan who did not think there was anything wrong with premarital sex but was still conflicted about whether or not she should. Eventually she decided not to have premarital sex. The researchers asked two different groups of participants, who had radically different views about whether refraining from premarital sex was a good thing, to read this vignette. On average, the participants they recruited from Washington Square Park in New York did not value refraining from having premarital sex, but those they recruited from a Mormon bible study did. When both groups were asked whether Susan *valued* refraining from having premarital sex, the participants from the Mormon bible study agreed that she did, while participants from Washington Square Park disagreed (Knobe & Roedder 2009).

Given the long-standing philosophical debate over happiness and the widespread impact of normative considerations on other folk concepts, a sustained empirical investigation into whether the ordinary concept of happiness is evaluative seems doubly promising.

IV. Investigating 'happiness'

We present six studies that attempt to gradually shed light on the ordinary understanding of happiness. The approach we take to studying the ordinary understanding of happiness involves presenting participants with cases that describe both a person's life and his or her psychological experiences. Then, participants are asked for their intuitive assessments of whether that person is happy. By systematically varying features of the person's life and psychological states, we investigate what is relevant to ordinary judgments of happiness.[15] Using this method, six studies provide evidence that the ordinary understanding of happiness reflects aspects of both evaluative and descriptive views. Similar to evaluative views, normative judgments have a substantive role in the ordinary understanding of happiness. Yet, similar to descriptive views, the ordinary understanding is focused on the person's psychological states and not the overall life they actually lived. Combining these two aspects, we offer a theory of ordinary happiness on which happiness depends on the *normative status* of an agent's *positive psychological states*.

Study 1: Normative evaluation influences judgments of happiness

This research began with a study in which participants were told about an agent named Richard. In one condition, Richard was described as a Red Cross field doctor who had both positive and negative mental states:

> Richard is a doctor working in a Red Cross field hospital, overseeing and carrying out medical treatment of victims of an ongoing war. He sometimes gets pleasure from this, but equally often the deaths and human suffering get

---

[15] Importantly, this method does not simply ask participants for their explicit theories of happiness (e.g., 'Could you be happy if you were not truly leading a good life?'). The approach we take has multiple benefits over this more explicit method. We take it that what is most relevant to the discussion of happiness is not people's opinion about theories about happiness, but rather how they ordinarily understand happiness. Moreover, we think there is very little reason to suppose that people have good epistemic access to the processes by which they form judgments of happiness. Thus, the approach we take allows us to derive a sense of what is relevant to the ordinary understanding of happiness even when people themselves may not be aware of it. In fact, we end up proposing a theory of the ordinary understanding of happiness that many people may actually find counterintuitive!

to him and upset him. However, Richard is convinced that this is an important and crucial thing he has to do. Richard therefore feels a strong sense of satisfaction and fulfillment when he thinks about what he is doing. He thinks that the people who are being killed or wounded in the war don't deserve to die, and that their well-being is of great importance. And so he wants to continue what he is doing even though he sometimes finds it very upsetting.

In the other condition, Richard was described as a Nazi death camp doctor who experienced the exact same type of mental states (differences in bold):

Richard is a doctor working in a **Nazi death camp,** overseeing and carrying out **executions and nonconsensual, painful medical experiments on human beings**. He sometimes gets pleasure from this, but equally often the deaths and human suffering get to him and upset him. However, Richard is convinced that this is an important and crucial thing that he has to do. Richard therefore feels a strong sense of satisfaction and fulfillment when he thinks about what he is doing. He thinks that the people who are being killed or **experimented on** don't deserve to **live,** and that their well-being is of **no** importance. And so he wants to continue what he is doing even though he sometimes finds it very upsetting.

After reading about one of these two agents, all participants were asked whether or not they agreed with the simple statement:
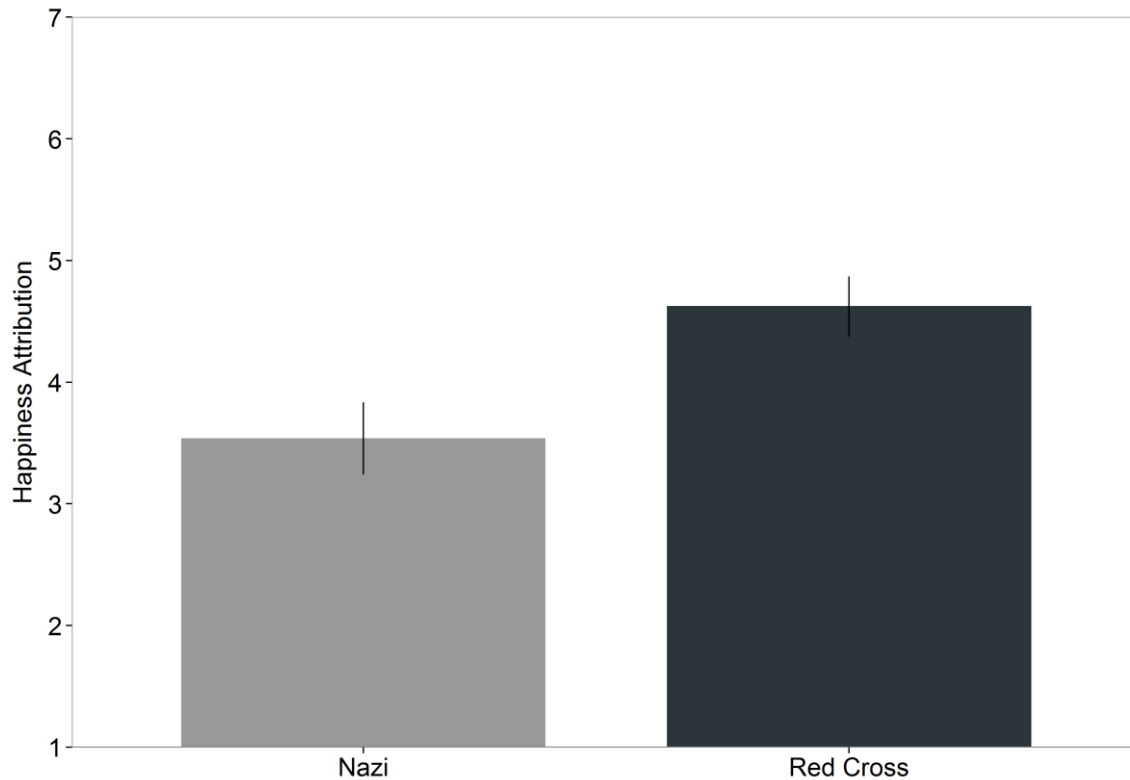
- Richard is happy.

Answers were recorded on a scale from 1 ('disagree') to 7 ('agree').

Comparing the two conditions, participants were significantly more likely to agree with the statement that Richard is happy when he was described as a Red Cross (Mean agreement: 4.63) doctor than when he was described as a Nazi doctor (Mean agreement: 3.54) (**Fig. 1**).[16]

---

[16] Participants were 182 undergraduates at UNC-Chapel Hill taking courses in philosophy, African-American studies, or sports and exercise science. Participants judged the agent to be happier in the good life (Red Cross) condition ($M = 4.63$, $SD = 1.217$) than in the bad life (Nazi) condition ($M = 3.54$, $SD = 1.486$), $t(180) = 5.4$, p $< .001$.

Figure 1. Agreement ratings for the statement 'Richard is happy' for the bad life condition (right) and good life condition (left). Error bars indicate bootstrapped 95% CI.

Although Richard was described as experiencing the same mental states in both conditions, participants overall tended to agree that the Red Cross doctor was happy, but that the Nazi doctor was *not*. On the face of it, this result undermines the widespread assumption that the ordinary understanding of happiness is only concerned with the mental states an agent experiences.

While we proposed that the difference in judgments of happiness was due to the difference in participants' moral evaluations, one might object that we have not yet provided evidence that it was specifically the moral evaluation of the agent that changed participants' judgments of happiness. Rather, the difference may have been due to one of the many descriptive differences in the two agents' lives. Based on the incomplete description of the doctors' mental states, participants could have reasonably inferred that the two doctors

actually experienced quite different sets of psychological states. If this is correct, then we may not yet have evidence that the ordinary understanding of happiness is genuinely evaluative.

To demonstrate that such differences in judgments of happiness actually result from the moral value of the agent's life, one would need to find a case in which some participants judge the life to be normatively bad and others judge it to be good, but the life they are judging is descriptively identical.

Study 2: Variance in normative evaluation predicts variance in attributions of happiness

In this study, all participants read a single vignette about an agent who experienced both positive and negative mental states. However, the vignette was constructed such that it was likely to elicit a wide range of moral evaluations about the agent. To achieve this, the vignette described a gay man named Bruce who, though conflicted because of a religious upbringing, began dating a man named Andrew.

> Bruce was brought up in a religious home and used to go to church every Sunday. However, since coming out as gay, Bruce has stopped going to church. He now meets new people at nightclubs, parties, and bars, instead of in church as he did before.

> One of Bruce's strongest desires is to find a man to have a long-term relationship with. He thinks that Andrew, whom he just recently met, might be what he is looking for. One of the reasons Bruce thinks this is that they are very compatible, sexually speaking.

> Occasionally Bruce feels a little guilty about being in a relationship with a man. It goes against his upbringing; he was raised to think that one of the most important things in life for a man is to find a wife. But, most of the time Bruce does not feel these kinds of guilty feelings. And, when he does, Bruce reminds himself that he thinks he really has no reason to. Bruce then also reminds himself that he really likes Andrew and the people they socialize with. And so he doesn't want to change anything about his lifestyle.

After reading the vignette, participants responded to three items in counterbalanced order:

- Bruce is happy. [Scale from 1 ('Agree') to 7 ('Disagree')]

- How much distress do you think Bruce feels? [Scale from 1 ('None at all') to 7 ('A lot')]

- Do you think Bruce's lifestyle is immoral? [Scale from 1 ('No') to 7 ('Yes')]

As expected, Bruce's lifestyle elicited a wide range of responses, with some of the participants judging Bruce's lifestyle to be immoral while others judged that it was not. Suggestively, participants' judgments of immorality were highly correlated with their judgments of happiness ($r = -.417$). By contrast, judgments of immorality were not at all correlated with judgments of distress ($r = .173$), suggesting that those who found Bruce's lifestyle to be immoral did not perceive him as feeling more distress.[17] More conclusively, when looking at how judgments of immorality and distress predicted judgments of happiness, we found that judgments of immorality predicted judgments of happiness in a way that was independent of perceived distress.[18]

This overall pattern can be more simply illustrated by dividing participants into two groups: (1) those who reported that Bruce's life was immoral and (2) those who reported that his life was *not* immoral. We can then compare the judgments made by the participants in these two groups. As predicted, participants who believed that Bruce was living an immoral life found him to be significantly less happy (Mean agreement: 3.83) than those who did not find his lifestyle to be immoral (Mean agreement: 5.02).[19] Moreover, we also compared how the two groups attributed distress to Bruce, to test if those who found his life

---

[17] Participants were 71 undergraduates at the University of Michigan. Judgments of immorality were highly correlated with happiness judgments, $r = -.417$, $p < .001$, but not with judgments of distress, $r = .173$, $p > .05$.

[18] The data were analyzed using a step-wise linear regression with judgments of immorality and distress entered as independent regressors and happiness judgments as the dependent variable. Unsurprisingly, judgments of distress by themselves predicted judgments of happiness ($\beta = -.481$, $p < .001$). More importantly, when judgments of immorality were included in the model with judgments of distress, judgments of immorality significantly predicted judgments of happiness $\beta = -.344$, $p = .001$, but we observed very little reduction in the predictiveness of judgments of distress ($\beta = -.421$, $p < .001$) in this second model. These results strongly suggest that the perception of immorality and distress account for independent variance in participants' judgments of happiness.

[19] The 5 participants who responded 'in between' to the question about whether Bruce's life was immoral were excluded from this analysis. A significant effect of morality on judgments of happiness was observed such that participants who responded that Bruce's lifestyle was not immoral judged him to be happier ($M = 5.021$, $SD = 1.246$) than those who responded that Bruce's lifestyle was immoral ($M = 3.833$, $SD = 1.505$), $t(64) = 3.26$, p < .002. We are only using a median split to highlight the patterns observed using correlation and regression analyses. None of our conclusions are drawn solely on the basis of these illustrative analyses.

to be immoral also attributed more distress. Yet, we found no significant difference in the groups' attributions of distress (**Fig. 2**).[20]
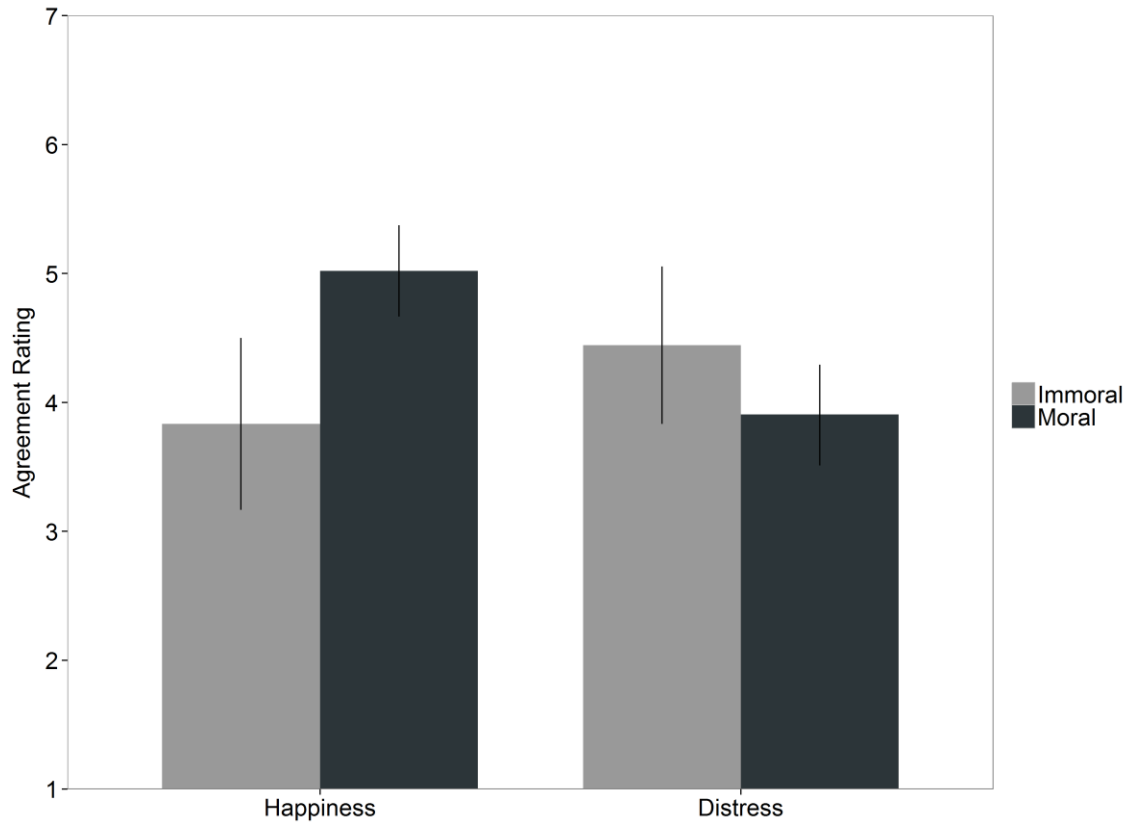
Figure 2. Attributions of happiness (left) and distress (right) for those participants who thought Bruce's life *was not* immoral (dark bars) and those who thought his life *was* immoral (light bars). Error bars indicate bootstrapped 95% CI.

In short, these results suggest that although both judgments of distress and judgments of immorality influenced attributions of happiness, the impact of normative judgments on judgments of happiness was *not* due to the difference in perceived distress. Participants' normative evaluations about Bruce's lifestyle were independently related to their judgments of happiness. This is particularly noteworthy given that all participants read

---

[20] $t(64) = -1.40, p = .165.$

exactly the same story about a single agent whose mental states were always described in exactly the same way. This result provides even better evidence than the previous study that the ordinary concept of happiness is partly evaluative and cannot be explained by perceived descriptive differences in the life the agent was living.

Additionally, this study helps to make clear a distinction between two different ways in which normative evaluations might influence judgments of happiness. On the one hand, some theories of happiness (e.g., life-satisfaction theories) have emphasized the importance of whether an agent's life lives up to *her own* evaluative standards. On this sort of view, an agent will not count as happy if her life does not measure up to her own evaluative judgments of what makes a life good. On the other hand, other theories (e.g., Foot 2001) have emphasized the importance of whether or not the life *actually is good*. On this sort of view, what an agent herself *thinks* is good may be totally irrelevant – she may simply be mistaken about what a good life actually is. What is relevant on this kind of view is only whether her life truly is good.

While these two evaluations often coincide, they can also easily come apart. In Study 2, all participants read the same vignette in which Bruce was described as living a life that he reflectively endorsed. Accordingly, if only the first kind of evaluative judgment (that which Bruce makes about his own life) were relevant to the ordinary understanding of happiness, then we should expect participants to uniformly judge that Bruce is happy. This is not what we observed. Instead, participants' judgments of happiness were substantially influenced by whether or not they judged Bruce's life was actually immoral.[21]

Study 3. The difference between happiness and unhappiness

One might still continue to press the objection that we have not yet provided sufficient evidence that the ordinary understanding of happiness is evaluative. In particular, one could object that we measured the perception of only one other psychological state:

---

[21] While it is possible that participants' moral judgments could have influenced how they perceived the agent's own values (and whether he was living up to those), we think it is unlikely that this could explain the pattern we observed for two reasons. (1) This explanation would suggest that participants who thought the agent's life was immoral should also believe that the agent was not satisfied with his life and should reflectively want to change it. However, the wording of the vignette specifically cuts against this interpretation, as the agent is described as not wanting to change anything about his life. (2) Even if participants ignored this portion of the vignette and did perceive the agent as differentially living up to his own values, then it is somewhat surprising that those who found his life to be immoral were not also more inclined to judge him as distressed, given that they should perceive the agent as much less satisfied with his life and as not living up to his own values.

distress. Even if it were true that participants who thought Bruce was living an immoral life didn't find him to be more distressed, they still could have perceived him to experience some other psychological state we didn't measure that actually explains why they judged he was less happy, e.g., dissatisfaction. If this is right, then there is no reason to assume the ordinary understanding of happiness is truly evaluative; we just haven't yet found the right descriptive psychological state to measure.

To go down this road would be unbearably tiresome. Suppose we do conduct another study and measure judgments of dissatisfaction only to find that moral evaluations don't influence judgments of dissatisfaction but do influence judgments of happiness. We are no better off than when we started. It's open to some other objector to press onward and suggest that we just haven't yet measured the relevant psychological state, which she might go on to suggest is actually pain or attitudinal displeasure or discontentedness or angst or ennui. Rather than entering into this tedious back and forth, it would be helpful to find a new way forward.

In a study previously reported in Phillips et al. (2011), participants read about an agent who was living a normatively good life, or an agent who was living a normatively bad life and were then asked to make either judgments of *happiness* or judgments of *unhappiness*.[22] More specifically, participants in the morally good condition were told about a woman named Maria who was working hard to raise her family, while participants in the morally bad condition were told about a woman named Maria who was doing anything and everything required to climb the social ladder in Los Angeles. After reading about one of these two agents, participants read a description of her mental states. In the happiness conditions, she was described as experiencing positive mental states, e.g., pleasure, satisfaction and enjoyment, and participants were asked the question:

- Do you think Maria is happy? [Scale from 'No' to 'Yes']

In the unhappiness conditions, she was described as experiencing negative mental states, e.g., displeasure, dissatisfaction and regret, and participants were asked the question:

---

[22] Even though this study has been reported previously, we summarize the results and experimental methods in detail for two related reasons. First, this study (Study 3) was originally published as an extension of Studies 1 and 2, and cited them as an unpublished manuscript by the second author. As such, the summary of these results here accurately represents the conceptual development of this research. Second, and relatedly, this study also forms a key piece of our theoretical argument and directly motivates the subsequent studies presented in this paper.

- Do you think Maria is unhappy? [Scale from 'No' to 'Yes']

Participants were assigned to one of four possible conditions: they could have either read about an agent living a morally good life or a morally bad life, and they could have been asked a question about happiness or a question about unhappiness.

The results of this study showed that while the normative value of Maria's life influenced participants' judgments of happiness, it did not influence their judgments of unhappiness. That is, participants tended to say that they thought Maria was happy when she was leading a morally good life, but not when she was leading a morally bad life. However, they equally thought Maria was unhappy, regardless of the type of life she was leading (**Fig. 3**).
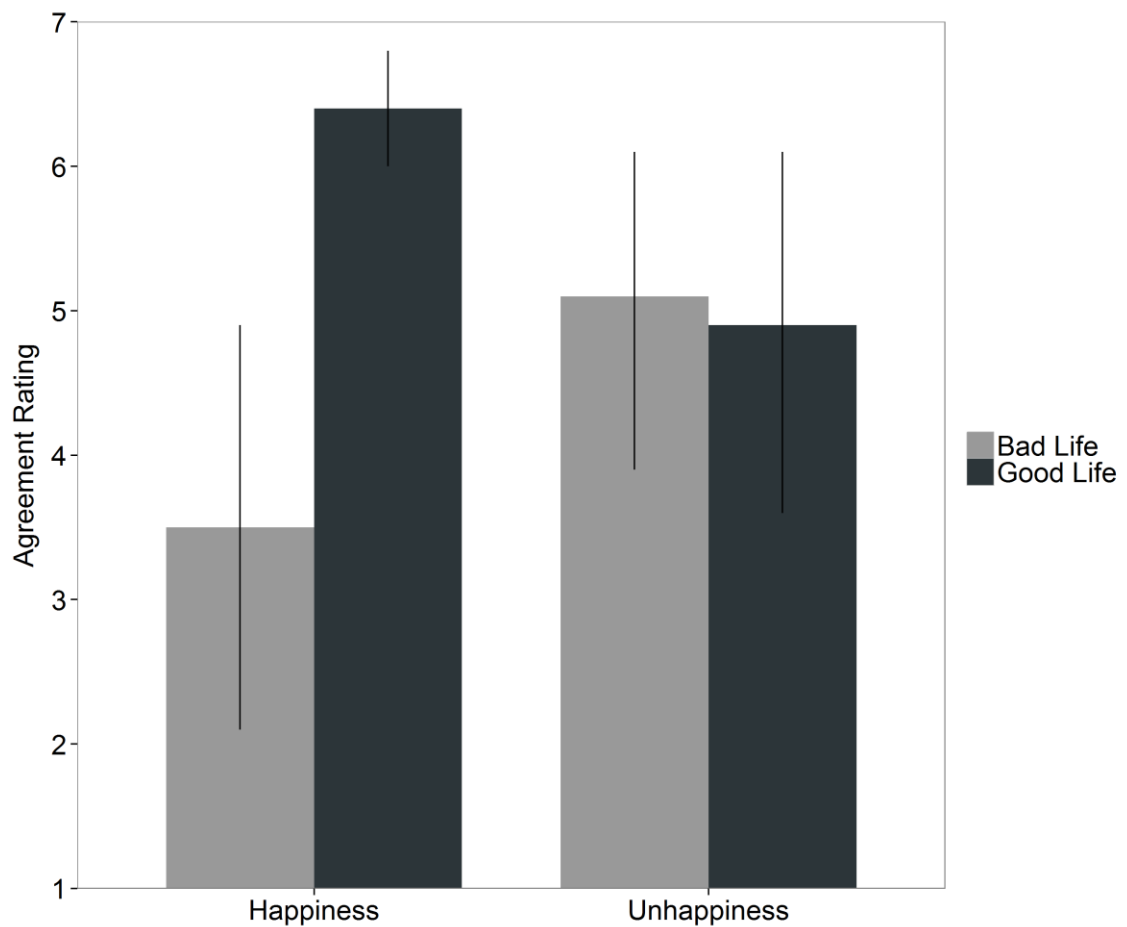
Figure 3. Attributions of happiness (left) and unhappiness (right) for the good life condition (dark bars) and the bad life condition (light bars). Error bars indicate bootstrapped 95% CI.

This approach to the study of the ordinary understanding of happiness is advantageous in that it allows us to preempt the sort of tedious back and forth that we sought to avoid. The results of this study are problematic for anyone who might have been tempted to argue that we did not specifically measure the relevant descriptive mental state in the previous studies, e.g., pain, displeasure, and so on. If the difference in happiness judgments in this study were due to a difference in the amount of displeasure, for example, the agent was perceived to have in the bad life condition versus the good life condition, then we should certainly expect that the difference in perceived displeasure would equally affect judgments of unhappiness. The same will be true of any other descriptive mental state that one might put forward as a possibility. In short, given that *only* judgments of happiness differed between the conditions, it seems unlikely that the difference in judgments of happiness is best explained by a perceived difference in the agent's descriptive mental states.

Study 4: Happiness, unhappiness, and negative mental states

A reasonable concern with the study discussed above (from Phillips et al. 2011) is that the happiness and unhappiness conditions differed both in the description of the mental states the agent experienced and in the question that participants were asked. That is, positive mental states are confounded with asking about happiness (in the happiness condition), negative mental states are confounded with asking about unhappiness and (in the unhappiness condition). Accordingly, it remains possible that the difference in the happiness condition but not in the unhappiness condition resulted not from the difference in the questions asked (happiness vs. unhappiness) but from the difference in the descriptions of the agents' mental states (positive vs. negative). If this turns out to be the case, this previous research may not reveal much about the ordinary understanding of happiness after all. To address this confound, we decided to describe the agent's mental states in exactly the same way in both the good life and the bad life conditions. More specifically, we tested whether the normative value of the agent's life would affect judgments of happiness, but not judgments of unhappiness, even when the agent only experienced negative mental states.

Participants in the good life condition read about a man named Mark who was volunteering in clinics in rural Africa, while participants in the bad life condition read about a man named Mark who was hopping from nightclub to nightclub across America. Regardless of which life Mark was living, he was always described as experiencing the same negative mental states:

> Day to day, Mark never feels satisfied. When he reflects on his life, he often feels distressed by the thought that he could do more, especially given his privileged upbringing. Each night, the images from the day haunt him to sleep. Still, each morning he decides to live as the day before, convinced that this is the life that he wants to live.

After reading about Mark, participants were asked to make simple judgment either about happiness or about unhappiness:

- Do you think Mark is happy?
- Do you think Mark is unhappy?

Participants responded to both questions on a scale from 1 ('Not at all') to 7 ('Very much so').

For the participants who made judgments of *unhappiness*, there was no significant difference between those who read about Mark living a normatively good life (Mean judgment: 4.96) and those who read about him living a normatively bad life (Mean judgment: 5.00). This is not surprising, given that Mark was always described as experiencing negative psychological states. However, participants did differ in their judgments of whether Mark was happy. Those who read about Mark living a good life judged him to be significantly more happy (Mean judgment: 2.68) than those who read about Mark living a bad life (Mean judgment: 1.63), even though he was always being described as experiencing the same negative psychological states (**Fig. 4**). The predicted interaction was significant.[23]

---

[23] Participants were 115 people (47 females, mean age 31) recruited from Amazon's Mechanical Turk. The data were analyzed using a 2 (normative value) x 2 (concept) ANOVA. We observed a main effect of Life Value $F_{(1, 111)} = 3.98$, $p < .05$, $\eta^2 = .04$, a main effect of Concept, $F_{(1, 111)} = 11.66$, $p = .001$, $\eta^2 = .10$, and, most importantly, a significant Life Value x Concept interaction effect, $F_{(1, 111)} = 4.58$, $p < .05$, $\eta^2 = .04$. The interaction was decomposed with planned comparisons, which revealed a significant difference in participants' responses to the question 'Do you think Mark is happy?' between the good life ($M = 2.677$, $SD = 1.536$) and bad life ($M = 1.63$, $SD = 1.079$) conditions, $t(56) = 3.0$, $p < .01$, but no significant difference in participants'
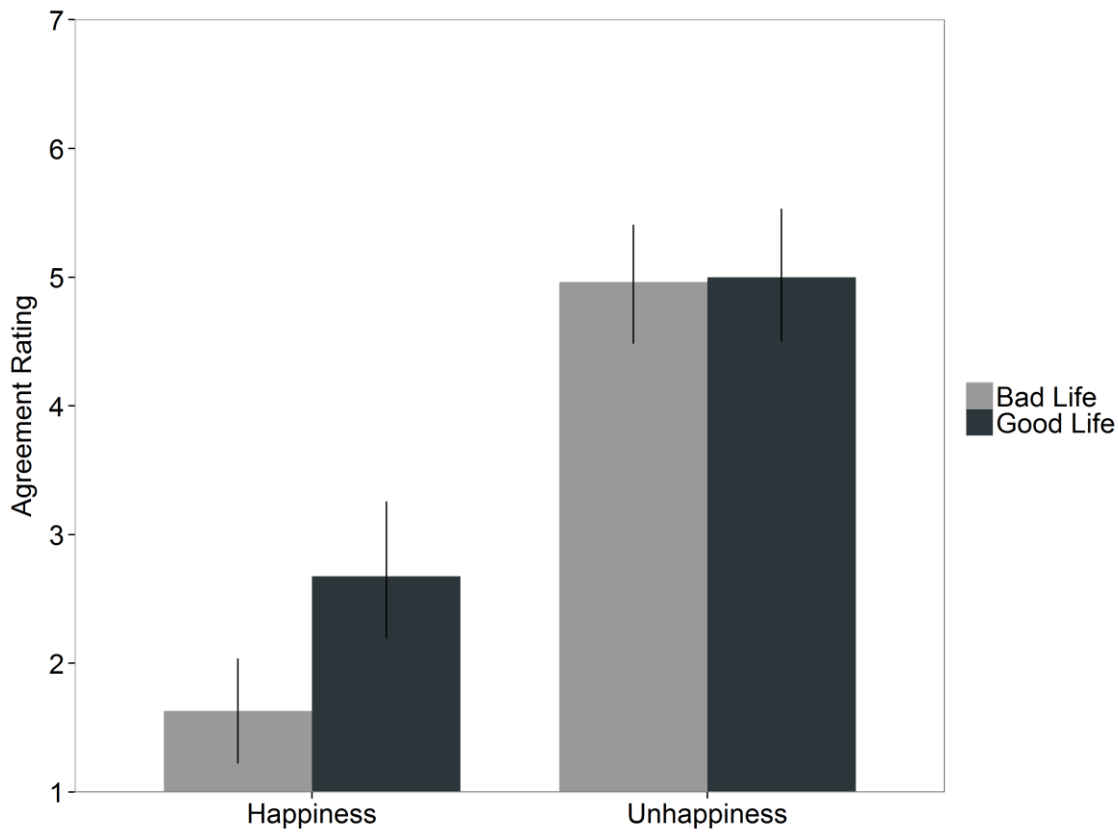
Figure 4. Participant responses to questions about happiness (left) and unhappiness (right) for the good life condition (dark bars) and the bad life condition (light bars). Error bars indicate bootstrapped 95% CI.

These results provide new evidence for a discrepancy between the way in which people ordinarily make judgments of happiness and unhappiness. When participants are trying to determine whether a given agent is happy, they seem to not only consider the mental state of that agent, but also take account of some normative evaluation of that agent. In contrast, when participants are determining whether a given agent is unhappy, this normative evaluation simply does not seem relevant.[24]

responses to the question 'Do you think Mark is unhappy?' in the good life ($M = 4.963$, $SD = 1.224$) and bad life ($M = 5.00$, $SD = 1.486$) conditions, $t < 1$.

[24] Notice that the key to our argument is the life value by concept interaction effect. In this study, and in all of the following ones, it happens that there was no significant difference in participant's judgments of unhappiness when reading about a person living a normatively good versus normatively bad life. While

Study 5: "Isn't happy" vs. "Is unhappy"

During a presentation of the previous studies, Stephen Stich pointed out that our way of accounting for this pattern of data makes a counterintuitive prediction. On our account, it seems that participants evaluative judgments should *not* affect their agreement ratings with statements of the form, '*S* is unhappy,' because this statement concerns 'unhappiness.' But at the same time, they *should* influence their agreement with statements of the form, '*S* isn't happy,' because this statement concerns 'happiness.' This should hold despite the striking similarity between both statements. Admittedly, we found the prediction counterintuitive, but we decided to take up Stich's challenge.

Participants once again read about an agent named Mark who experienced negative mental states (distress, dissatisfaction, etc.) and he was either described as living a morally good life (volunteering in rural Africa) or described as living a morally bad life (hopping from nightclub to nightclub across America). After reading about the agent's life, participants were asked whether they agreed or disagreed with one of two statements:

- Mark isn't happy.
- Mark is unhappy.

Participants rated their agreement on a scale from 1 ('Completely Disagree') to 7 ('Completely Agree'). Finally, to continue to control for whatever uninteresting differences in perceived mental states there may be, participants also rated how much distress they believed Mark experienced.

Despite the minimal difference between the statements, we once again observed the predicted asymmetry between judgments of happiness and judgments of unhappiness (**Fig. 5**). Those who rated their agreement with the statement 'Mark is unhappy' were equally likely to agree with the statement, regardless of the type of life that Mark was described as living a good life (Mean agreement: 5.73) or bad life (Mean agreement: 5.80). However, there was a significant difference in participants' agreement ratings for the statement 'Mark isn't happy.' As predicted, those who read about Mark living a normatively good life agreed less that he

---

interesting (and rhetorically convenient), this lack of a difference is not particularly important for our point. What is crucial for the point we want to make is for normative evaluations to have *more* of an effect on judgments of happiness than they do on judgments of unhappiness. Thus, even if future research uncovers cases in which participants' judgments of unhappiness also differ between conditions, our point will still stand as long has judgments of happiness are significantly more influenced.

'isn't happy' (Mean agreement: 5.25) than those who read about Mark living a normatively bad life (Mean agreement: 6.47). The predicted interaction was significant even though we statistically controlled for the amount of distress participants perceived Mark as experiencing.[25]

---
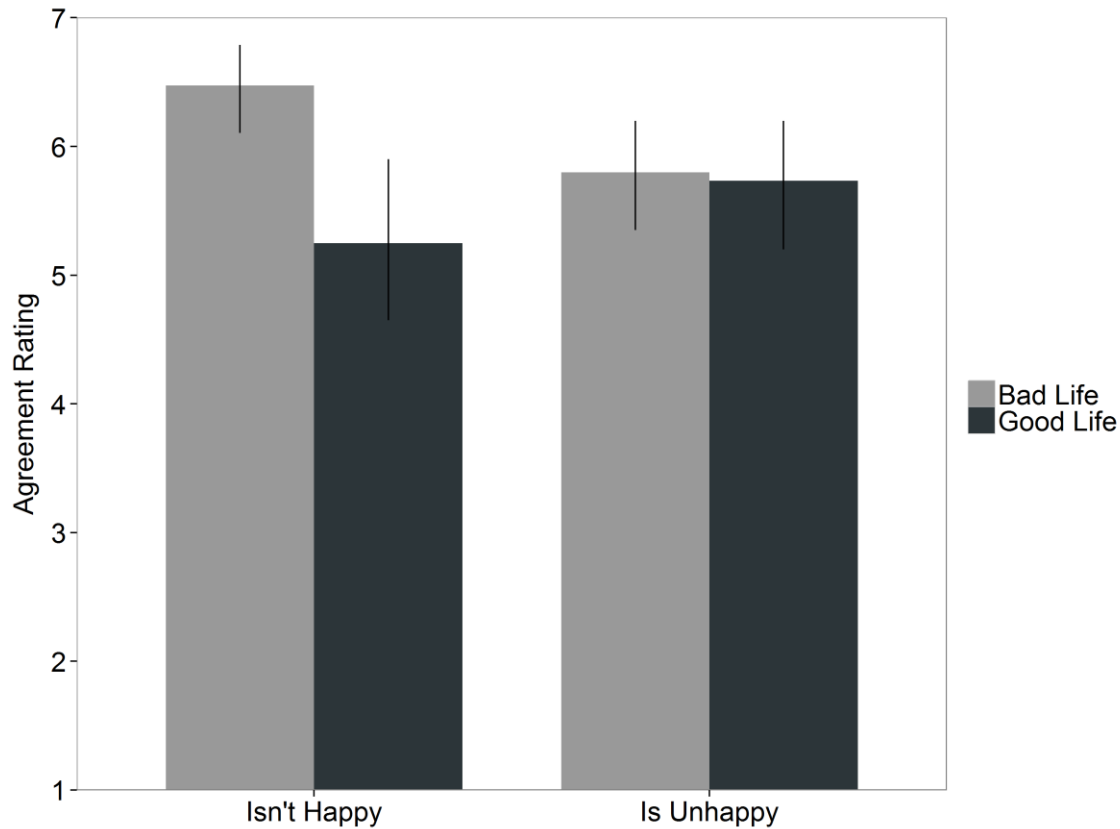
[25] Participants were 70 people (47 females, mean age 37) spending time on Amazon's Mechanical Turk website. Four participants were excluded for failing a simple control question. The data were analyzed using a 2 (Life Value) x 2 (Concept) ANCOVA with judgments of distress entered as a covariate. We observed a main effect of life value, $F(1, 65) = 4.96$, $p < .05$, $\eta^2 = .07$, no main effect of concept, $p > .05$, and no main effect of perceived distress, $p > .05$. Most importantly, there was a significant Life Value x Concept interaction effect $F(1, 65) = 4.12$, $p < .05$, $\eta^2 = .06$. This interaction was decomposed with planned comparisons, revealing that participants agreed more that 'Mark isn't happy' in the bad life condition ($M = 6.41$, $SD = 0.795$) than in the good life condition ($M = 5.26$, $SD = 1.522$), $t(34) = 2.79$, $p < .01$. By contrast, there was no similar difference in their judgments of unhappiness between the bad life condition ($M = 5.84$, $SD = 1.015$) and the good life condition ($M = 5.73$, $SD = 1.033$), $t < 1$. Including the four participants who failed the control question in this analysis only increased the strength of the observed effects.

Figure 5. Participants' agreement ratings with the statement about happiness (left) and unhappiness (right) for the good life condition (dark bars) and the bad life condition (light bars). Error bars indicate bootstrapped 95% CI.

The results from the previous five studies have shown a highly consistent pattern: when people ordinarily try to figure out whether or not some agent is happy, their judgments are influenced not only by whether the person experiences positive or negative psychological states, but also by their normative evaluations of that person.

Study 6: Toward a normative psychological theory of happiness

While ordinary judgments of happiness have been shown to be influenced by normative evaluations, a further question remains about exactly which normative judgments are relevant. One possibility is that the relevant normative judgments, as in evaluative theories of happiness, primarily concern how the agent's life is *actually* going, in a way that is completely independent of what the person thinks or believes. Another possibility is that the relevant normative judgments are about the life that the person *believes or understands* herself to be living.

The key distinction between these two possibilities can be seen in the example Richard Kraut offered about a man who wants most to be loved, admired and respected by his friends (1979). His 'friends' collude in an elaborate plot to give him every reason to believe that he is loved, admired and respected by them, though this is not actually the case. Unsurprisingly, the man is taken by the deception. In this case, we can imagine two separate 'lives' that we might make normative judgments about. Like Richard Kraut suggests, we may judge that the life that the man is actually living is straightforwardly not going well – it is not a good life. This normative judgment is the one that evaluative theories of happiness have focused on. Yet, independent of how the man's life is actually going, we may also judge that the life that the man understands himself to be living *is* going well – *that life* is a good life. These two normative judgments can clearly vary independently of one another.

This same distinction may also be applied to the question of whether one is living a *morally* good or bad life. It is certainly possible, for example, for one to take oneself to be living a life that is leading others to be much better off, but it turn out to be the case that one

is actually living a life that is leading others to be significantly worse off. With this distinction in mind, we can ask whether the relevant normative judgments in the previous studies were evaluations of the life that the agent was *actually* living, external to the agent's understanding, or whether the relevant normative judgments were evaluations of the life that the agent understood herself to be living. As the previous studies do not distinguish between these two possibilities, we sought to pull apart them apart with one final study involving cases in which an agent may be understandably misled about how her life is going.

Participants read about a nurse named Sarah who worked at a children's hospital and whose mental states were always described as highly positive.

> 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.

However, we varied both whether Sarah understood herself to be living a life that led others to be better off and whether the life she was living actually led others to be better off. Accordingly, Sarah was described as understanding herself to be living a life that was good for others in half of the cases:

> The reason Sarah feels this way is that she tries to help the sick children by giving them medicine that tastes like candy.

While in the other half of the cases, Sarah was described as understanding herself to be living a life that was leading others to be significantly worse off:

> The reason Sarah feels this way is that she tries to poison the sick children by giving them pesticides that taste like candy.

Along the other dimension, the life that Sarah was actually living did lead others to be significantly better off in half of the cases:

Ten years after Sarah died, a new medical discovery was made which

suggested that the pesticides [medicine] had made some of the

children less sick.

While in the other half of the cases, the life that Sarah was actually living led others to be significantly worse off:

Ten years after Sarah died, a new medical discovery was made which

suggested that the medicine [pesticide] had made some of the

children more sick.

After reading one of the four possible vignettes, participants were asked whether they agreed or disagreed with a single statement about happiness:

- Sarah was happy.

Participants responded on a scale from 1 ('Completely Disagree') to 7 ('Completely Agree').

Subsequently, participants were asked a series of control questions to make sure that they were not incorrectly imagining that Sarah actually experienced negative psychological states. These questions ensured that participants believed Sarah did experience the three components of the psychological definition of happiness: high levels of positive affect, low levels of negative affect, and high life satisfaction (Diener, Scollon, & Lucas 2003).

The design of this study allowed us to test whether participants' judgments of happiness were affected by the value of the life the agent understood herself to be living or by the value of the life the agent was actually living. Participants' judgments of happiness revealed that they were significantly affected by the value of the life the agent took herself to be living, but not by the value of the life the agent was actually living. Accordingly, participants rated Sarah as happier when she understood herself to be living a life that was improving the children's wellbeing (Mean agreement: 6.67) than when she understood herself to be living a life that was causing the children to be harmed (Mean agreement: 5.83). By contrast, there was no significant difference in participants' happiness ratings based on whether the life Sarah led actually did improve the wellbeing of children (Mean agreement:

5.96) or instead significantly harmed their wellbeing (Mean agreement: 5.83) (**Fig. 6, left panel**).[26]

To forestall any concern that this design did not allow us to adequately vary participant's normative judgments about both the agent's actual life and the life the agent understood herself to be living, we asked an additional group of participants to read the same vignettes and rate their agreement with two statements that corresponded to the two relevant normative judgments.[27]

- Ignoring everything that Sarah thought about the life she was living and considering only what actually happened, Sarah was living a good life.
- Considering only what Sarah thought about the life she was living and ignoring everything that actually happened, Sarah was living a good life.

Considering first the normative judgments of the agent's actual life, participants' responses indicated that their judgments were affected by whether Sarah *understood* herself to be helping or harming children, such that they more agreed that she was actually living a good life when she took herself to be helping children (Mean agreement: 5.23) than when she took herself to be harming children (Mean agreement: 1.89). While not specifically intended, this is not particularly surprising. The relevant question, however, is whether participants' normative judgments about Sarah's actual life were adequately impacted by whether Sarah was *actually* helping children. Considering the overall pattern, participants' normative judgments did, in fact, appear sensitive to whether Sarah was actually harming or helping children. However, this was only the case when Sarah understood herself to be helping the children.

More specifically, when Sarah understood herself to be poisoning children, participants did not think that having happened to help them significantly made her life any

better.[28] By contrast, when Sarah understood herself to be helping children, participants agreed that Sarah's life was significantly better when she actually was helping children (Mean agreement: 6.00) than when she was actually harming them (Mean agreement: 4.57) (**Fig. 6, right panel**).[29] This pattern suggests that, at least in the conditions in which Sarah understood herself to be helping children, we also successfully change participants' normative judgments of Sarah's actual life. Importantly, however, this difference in the value of Sarah's *actual* life was not mirrored in participants' judgments of whether Sarah was happy in those same conditions.

Focusing next on the question about the normative value of the life that Sarah understood herself to be living, participants' judgments indicated that they did think that the life Sarah understood herself to be living was significantly better when she took herself to be giving children medicine (Mean agreement: 6.33) than when she took herself to be poisoning them (Mean agreement: 5.07). Moreover, participants' judgments were not affected by whether or not Sarah actually was helping children or was actually harming them.[30] This suggests that we were able to adequately vary the value of the life Sarah understood herself to be living. When considered as a whole, the pattern of participants' normative judgments about the life the agent understood herself to be living bears a striking resemblance to previous participants' judgments of happiness (**Fig. 6, middle panel**).

---

[28] One likely explanation of this is simply that participants ratings were so low on the available agreement scale when Sarah understood herself to be poisoning children that it was simply not possible to detect an effect of actual life value. This is often referred to as a 'floor effect.'

[29] Participants were 105 people (61 women, mean age 30.42 years, $SD = 8.61$) recruited from Amazon's Mechanical Turk. 21 participants were excluded for either missing the previously discussed control questions or taking less than 20 seconds to read the vignette (average completion time was over 60 seconds). The data were again analyzed using a 2 (understood life value: good vs. bad) x 2 (actual life value: good vs. bad) ANOVA. We observed a main effect of understood life value, $F(1, 80) = 86.67$, $p<.001$, $\eta^2 =.522$. In addition, observed a trend of actual life value, $F(1, 80) = 2.68$, $p=.105$, $\eta^2 =.032$, and a significant interaction effect $F(1, 80) = 4.75$, $p=.032$, $\eta^2 =.056$. Planned comparisons investigated the interaction effect. When Sarah understood herself to be helping children, participants judged that her life was significantly better when she actually was helping children ($M = 6.00$, $SD = 0.97$) than when she was actually harming children ($M = 4.57$, $SD = 1.66$) $t(32.935) = -3.33$, $p = .002$. However, when Sarah understood herself to be harming children, participants evaluations of her life did not significantly change if she was actually helping children ($M = 1.82$, $SD = 1.76$) or harming children ($M = 1.96$, $SD = 1.89$), $p = .80$. Including the excluded participants in the analysis did not significantly change these effects.

[30] The data were again analyzed using a 2 (understood life value: good vs. bad) x 2 (actual life value: good vs. bad) ANOVA. We observed a main effect of understood life value, $F(1, 80) = 12.63$, $p<.001$, $\eta^2 =.136$, but no effect of actual life value, $F < 1$, and no interaction effect, $F < 1$. Planned comparisons were used to investigate the effect of understood life value which revealed that participants agreed that the life Sarah understood herself to be living was better when she took herself to be helping children ($M = 6.33$, $SD = 0.81$) than when she took herself to be poisoning them ($M = 5.07$, $SD = 2.07$), $t(59) = -3.78$, $p < .001$. Including the excluded participants in the analysis did not change the observed effects.
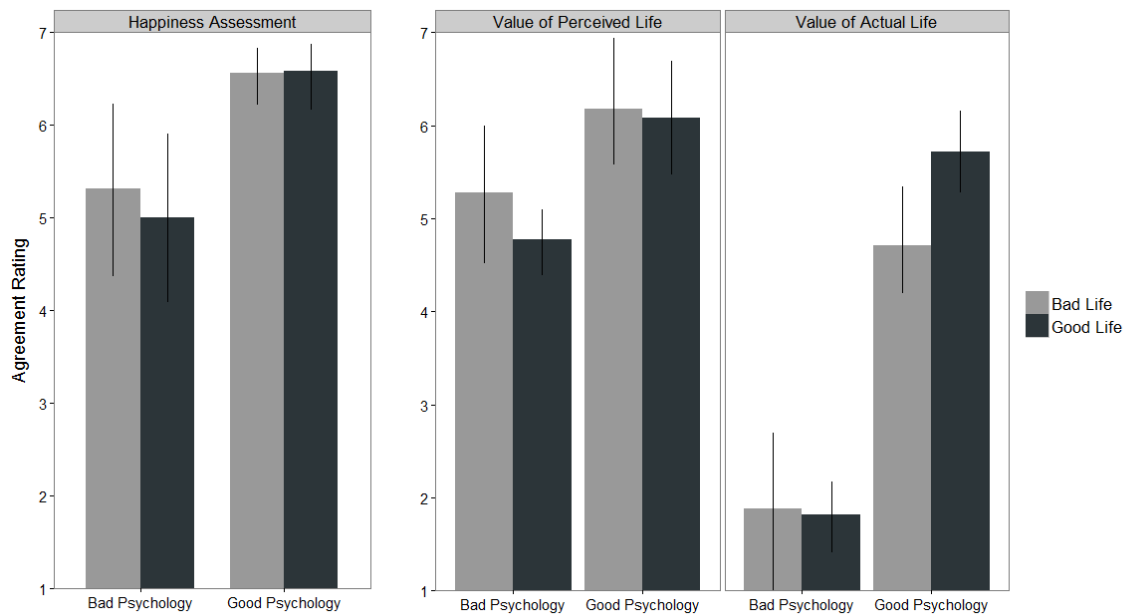
Figure 6. Participants' agreement ratings with the statement about happiness, value of perceived life value, and value of actual life for an agent who understood herself to be helping children (left) or hurting them (right), and who actually was helping children (dark bars) or hurting them (light bars). Error bars indicate bootstrapped 95% CI.

The key question, which we now have the ability to answer, is which of these two normative judgments is relevant for ordinary happiness. Are ordinary judgments of happiness sensitive to the value of the life the person *actually* lives, or are they sensitive to the value of the life the person *understands herself to be living*? The data from this study suggest that it is the value of the life the person understands herself to be living that is central to ordinary happiness.

V. General Discussion

The six studies presented in this paper provide evidence that normative value is relevant to ordinary happiness, and, more specifically, that the relevant normative judgments focus on the life the agent understands herself to be living. First, we consider one alternative explanation for the patterns we observed. Then, we take up the task of outlining the sort of theory which could adequately capture the way that people ordinarily make judgments of

happiness. Finally, we step back and discuss the philosophical and psychological implications of these findings.

The multiple senses objection

One way of objecting to the claim that ordinary judgments of happiness are essentially evaluative would be to suggest that the term 'happiness' actually has multiple senses, one of which is descriptive (similar to 'pleasure') and one of which is evaluative (similar to 'well-being'). Indeed, there has been a large amount scholarly speculation that this is the case (Feldman 2010; Foot 2001; Haybron 2000; Kraut 1979; Sizer 2010). According to this objection, there is no single understanding of happiness which focuses on the agent's psychological states but is also evaluative. Rather, these two aspects represent features of two independent senses of 'happiness'. Accordingly, combining them into a single, unified understanding of happiness would be mistaken.

While we do not doubt that people sometimes use the term 'happy' in a way that seems synonymous with pleasure (e.g., 'I'm very happy to meet you.') or with well-being (e.g., 'I wish you a happy life.'), we think there is good reason to suspect that this could not adequately account for the full pattern of data. One major obstacle is that this objection cannot explain why we observed a difference in judgments about happiness *but not* judgments about unhappiness. Thus, even if the multiple senses objection could account for participants' judgments of happiness, it would also need to explain why these independent senses were not at play in judgments of unhappiness. After all, we use 'unhappy' in a ways that seem synonymous with displeasure (e.g., 'Sad music makes me unhappy.') and with a lack of well-being (e.g., 'Their marriage has always been unhappy.').[31]

Even more conclusively, recent research (Knobe, Prasada & Newman *forthcoming*) has developed tests for concepts that have a 'dual-character' such that a single concept can be used in both normative and descriptive senses (e.g., *mother*, *rock music*, *criminal*). Employing these tests, researchers have already found evidence that happiness does *not* fit in the

---

[31] This objection also makes an empirical prediction concerning the distribution of participants' responses in the cases in which an agent was described as having positive psychological states but a normatively bad life. Specifically, it suggests that the distributions of participants' responses should be bimodal, indicating that some participants used the purely descriptive sense (thus agreeing that the agent was happy) while others employed a purely evaluative sense (thus disagreeing that the agent was happy), but that very few participants would mix up these two senses. However, the data simply do not bear out this prediction: the distributions of responses were consistently unimodal in these exact conditions. These data are available upon request from the first author.

category of dual-character concepts. Happiness no more admitted of a dual character than emotions such as sadness, anger, fear or disgust (Chituc 2012). This finding directly contradicts what we would expect if happiness genuinely had distinct normative and descriptive senses. Thus, we find it unlikely that an objection based on multiple senses can explain these results.

The good in happiness

Even if one accepts that there is a single evaluative understanding of happiness, one might still be at a loss as to how we should make sense of the ordinary understanding of happiness. While it is highly likely that any specific theory proposed at this still early stage will be overturned upon the discovery of new evidence or further theorizing, we propose one way of accounting for ordinary happiness.

Let us begin by considering how the present studies relate to the long-standing debate over happiness. On descriptive theories of happiness, whether or not an agent is happy is a matter of figuring out exactly which psychological states that agent experiences and deciding which ones constitute happiness. In contrast, evaluative theories of happiness hold that if an agent is not living normatively good life, then she won't truly be happy no matter which psychological states she experiences. When contrasting the two sides of this debate, it may be tempting to take the results of these studies to be evidence *against* the descriptive theories and *for* evaluative theories. However, this would be too quick. These two groups of theories differ along two distinct dimensions. First, they differ in whether or not the normative value of an agent's life is relevant to whether the agent is happy. Second, they also differ in what they focus on when deciding whether someone is happy. Descriptive theories tend to focus primarily on the *psychological states* that an agent experienced, while evaluative theories tend to focus primarily on the *actual life* the agent has led (**Fig. 8**).

	Psychological States	Overall Life
Normative Value	?	Evaluative Theories, e.g., Foot; Annas

	Descriptive	
No Normative Value	Theories, e.g., Feldman; Diener	X

Figure 8. Normative value by focus matrix of possible theories of happiness.

The ordinary understanding of happiness seems to deviate in important ways from both evaluative and descriptive theories of happiness. Unlike descriptive theories, normative information about the agent is central to the ordinary understanding. Accordingly, it is unlikely that ordinary happiness is merely a matter of possessing some particular psychological state or set of psychological states (bottom row). Yet, unlike evaluative theories of happiness, the ordinary understanding of happiness does not focus on the overall life the agent actually lived. Rather, it focuses on the life the agent understands herself to be living, which is quite clearly a matter that is solely determined by the psychological states the agent experiences. Thus it also seems unlikely that ordinary happiness is simply a matter of having led a good life. Instead, ordinary happiness focuses on the psychological states of agents, but does so in a genuinely normative way. Ordinary happiness seems to be a matter of possessing positive psychological states that are normatively right to experience, given the life one understands oneself to be living.[32]

It will be important to be more precise about the way in which a psychological state may be right or wrong to possess. On our proposal, a positive psychological state is wrong to possess if one ought not be in that state, given the life one understands oneself to be living. Clearly, one ought not experience satisfaction with one's life if one understands that life to be one that directly harms children's wellbeing by poisoning them. Positive psychological states that are not wrong to possess count (in the relevant sense) as right to possess. Thus, to say that someone is happy would be to say that she experiences positive

---

[32] Important questions clearly remain about what is required for an agent to understand herself to be living a particular type of life. Such an understanding will presumably require more than merely believing that one is living a morally good life - even when one has sufficient reason to believe that one is actually not living that life. Yet, rather than trying to offer some specific account which is unlikely to be correct, we wish to be intentionally imprecise about this issue, and hope that this question will be taken up in future work on this topic.

states like pleasure or satisfaction, and that given the life she understands herself to be living, it is not the case that she ought not experience those states.[33]

It should not be too difficult to see how this theory can account for participants' judgments. Consider the case of the nurse in Study 6 who derives pleasure and satisfaction from trying to save children's lives. It seems obvious that there is no reason that the nurse ought not experience satisfaction given that she understands herself to be living a life in which she makes children's lives better. In contrast, it is clear that the nurse ought not experience pleasure or satisfaction when she understands herself to be living a life in which she is harming children by poisoning them. This too remains the case whether or not she is actually successful in causing the children harm.

Before continuing on, it may be helpful to end by explicitly specifying the sort of normative evaluations that we think are at play in ordinary judgments of happiness. On our view, it is specifically *moral* normativity that is relevant to happiness and not, by contrast, the normativity of practical rationality or prudence. More precisely, then, to be happy in the ordinary way is to experience positive psychological states when it is *not* the case that one *morally ought not* have them, given the life one understands oneself to be living. To see why it is specifically moral normativity matters for the proposed theory of ordinary happiness, consider again the much-discussed case of the Nazi war criminal who experiences pleasure and contentment while living on an Argentinian beach. Suppose that he has the goal of not becoming depressed and committing suicide, which he fears could easily happen. According to practical rationality then, he *ought* to experience the pleasure and contentment he does when living on the Argentinian beach. After all, he intends to not commit suicide and believes that feeling good will help him achieve his intended goal. Thus, if we are concerned with the normativity of practical rationality or prudence, the Nazi war criminal should count as happy, as it is clearly *not* the case that he ought not experience the pleasure and contentment he does. By contrast, one certainly might think that he *morally ought not* to experience those positive states given his understanding of his life. On our account of

---

[33] One might be tempted by the simpler formulation according which happiness is a matter of possessing positive mental states that one *ought* to possess. However, this simpler formulation is overly strict as it rules out amoral cases in which it is neither the case that one ought to possess the mental state nor the case that one ought not to possess them. The pleasure and satisfaction we have in such amoral cases, e.g., when eating ice cream, presumably should count toward our happiness, so we prefer the less restrictive formulation.

ordinary happiness, as long as one holds that he ought not experience those states, the Nazi would not count as truly happy.

While the tentative theory we have offered can account for the ordinary judgments of happiness in the six studies we presented, future studies will certainly be what decides between our proposed theory and other alternatives ways of accounting for the ordinary understanding of happiness. Yet regardless of which account is eventually accepted, we take these studies to provide sufficient evidence that any adequate account of ordinary happiness will have to provide some role for normative judgment.

Finally, it will be informative to contrast this proposal with other theories that give normative judgments an importantly different role. Recall, as discussed in § II, that descriptive theories of happiness often involve a particular sort of normative judgment. Similar to some theories of well-being (e.g., Tiberius 2008; Tiberisu & Plakias 2010), life-satisfaction theories of happiness hold that, to be happy, agents must be reflectively satisfied with their own lives (Kekes 1982; Suikkanen 2011). While such judgments of life-satisfaction clearly involve normative evaluations, such theories are still purely descriptive because the relevant normative judgments are simply mental states the agent experiences. To know whether or not someone is happy, all one would need is an accurate list of all the psychological states the person experiences, even if some of the states on this list involve the agent herself forming a normative judgment. In contrast, the evidence we have provided suggests that ordinary happiness genuinely evaluative in that the relevant normative judgments are about whether the agent *ought* to be experiencing the positive mental states she experiences. We simply cannot know whether an agent is happy by referencing an accurate list of her mental states – we also need to know whether she ought to be experiencing those states, and this is simply not a descriptive question.

Philosophical and psychological implications

Even if one were convinced that we have accurately accounted for the ordinary understanding of happiness, one might still wonder why either philosophers or psychologists should be interested in this ordinary understanding. While we think this discovery is relevant for both philosophical and the psychological research, we do not think it is relevant both research projects in the same way. We take up the relevance for these two disciplines in turn.

Ordinary happiness and ethics

One response to this line of research would be to be skeptical about its relevance for the philosophical discussion. Philosophers, after all, are presumably not interested in what people *think* happiness is, but rather in what happiness *actually* is. In making this point clear, one might point out that we've long known that the ordinary understanding of physics differs dramatically from philosophers' discussions of it (McCloskey & Kohl 1983; McCloskey, Washburn & Felch 1983), but that this divergence clearly does not give us reason to revise *that* philosophical discussion! We simply could not more agree with this conclusion, in the case of physics. What is not clear, however, is whether this is an appropriate analogy.

The irrelevance of the ordinary understanding of physics follows straightforwardly from considering the evidential relation between the ordinary understanding and the underlying physics itself, which is what philosophers are concerned with. In that case, there is a relatively weak evidential relation between the ordinary understanding and the object of philosophical discussion, and even if there is *some* evidential relation, it is certainly not as strong as the evidential relation between rigorous scientific experiments and the object of philosophical interest. Accordingly, it is not surprising that contemporary philosophers simply do not show an interest in the ordinary understanding of physics, since this ordinary understanding is a comparatively terrible way of accessing physics itself.

By contrast, philosophers who study happiness *do* seem to show a great deal of interest in the ordinary understanding of happiness. It is instructive to consider why this might be. Given that the philosophical discussion of happiness is specifically concerned with happiness *itself* (just as much as philosophy of physics is concerned with physics itself), one obvious reason may be that philosophers take the ordinary understanding of happiness to have some nontrivial epistemic relation to happiness. Of course, nothing demands that the ordinary understanding have such a strong evidential relation, and perhaps happiness will turn out like to be like physics. However, until we are presented with better evidence that this is the case, we are content to throw our lot in with the many other philosophers who also seem to think that the ordinary understanding is relevant to the philosophical discussion in the case of happiness.[34] Accordingly, we maintain that, all else being equal, these studies

---

[34] Note that nothing we have said suggests that the ordinary conception is the *only* thing that is relevant to the philosophical discussion. Other concerns, e.g., consistency, should perhaps be weighted even more heavily than

should inform our theorizing about happiness by providing *some* evidence about happiness itself.[35]

Yet, even accepting that the ordinary understanding of happiness may be broadly relevant to the philosophical discourse, one may be left wondering what conclusions should be drawn from the present research. We hope this research will inspire further discussion on this exact question, but we take up two specific implications.

The most straightforward implication is that, if our proposed theory of the ordinary understanding of happiness turns out to be right, then it provides evidence that the study of happiness may turn out to partly be a branch of normative ethics. Whether somebody is happy, in other words, depends in part on normative questions about what psychological states ought not be had under what understood conditions.

Relatedly, a second implication about the study of happiness is that thoroughgoing research on happiness cannot remain completely neutral on the difficult questions discussed within the field of meta-ethics. To see this, consider cases in which people's moral disagreements lead them to disagree about whether somebody is happy (e.g., Bruce in Study 2). Meta-ethical theories according to which there are objective facts about what moral convictions are true imply that some of the participants in our study were simply wrong. These theories imply that there is an objective fact about whether Bruce is happy, a fact partly settled by some wholly objective moral fact about whether Bruce's psychological states are the ones he ought to have. By contrast, meta-ethical theories according to which moral facts are subjective, or speaker- or group-relative, imply that we cannot sensibly talk about Bruce's happiness without first relativizing our claims to the norms of particular or individuals or groups. Accordingly, participants may not actually be disagreeing with each other when they utter seemingly contradictory statements about Bruce's happiness.

Notice that the way that meta-ethical issues bear on Bruce's happiness is indirect. It

---

the ordinary understanding in theorizing about happiness. After all, if the ordinary understanding of happiness turns out to be internally inconsistent, one might take this to be good evidence that it does not bear a strong evidential relation to happiness.

[35] Another reason happiness researchers and philosophers may be concerned with the ordinary understanding of happiness is that people ordinarily care an awful lot about (what they think of as) happiness. In fact, we imagine that this is one of the very reasons that there has been so much interest in studying happiness. To the extent that happiness researchers or philosophers abandon all interest in happiness-as-it-is-ordinarily-understood in favor of some purely technical notion of happiness, call it happiness*, they run the risk of abandoning the reason this research was motivated in the first place. At the very least, these researchers would now need to provide some independent motivation for caring about happiness* rather than happiness as it is ordinarily understood.

is not the case that Bruce's happiness depends in any direct way on meta-ethical issues; rather, the point here is that whether it is *objectively true* that Bruce is happy depends directly on meta-ethical issues. By contrast, questions of whether or not Bruce is happy directly depend on issues in *normative* ethics.[36]

Ordinary happiness and affective science

　　The studies we have discussed here reveal that the ordinary understanding of happiness combines both descriptive information about an agent's psychological states and evaluative judgments about whether the agent ought to possess those states. By contrast, the scientific definition of happiness focuses solely on descriptive information about the agent's psychological states, i.e., some combination of high positive affect, low negative affect, and high life satisfaction (Diener, Scollon, & Lucas 2003; Lyubomirsky, King & Diener 2005).

　　One important implication of this difference between the scientific definition and the ordinary way of understanding happiness is that it indicates that the methods employed by affective scientists cannot investigate the entirety of happiness, as it is ordinarily understood, because this is in part a normative question.

　　This, of course, does not at all suggest that affective scientists should revise their definition to more closely align with the ordinary understanding of happiness. Rather, it suggests that affective scientists and positive psychologists should continue to pursue their research, but with the understanding that the ordinary understanding of happiness differs in an important respect. On this picture, the present research is best understood as a necessary complement to the research in affective science.

VI. Conclusion

The study of happiness, from ancient history to contemporary research, has continually presented two different ways of thinking about the topic. On one side are evaluative views of happiness, which hold that being happy requires one to live a good life. On the other side

---

[36] This point can be straightforwardly generalized to the superset of findings that illustrate the role of moral judgment in the ordinary understanding of other concepts, e.g., intentional action, causation, or freedom. However this generalization is conditional on the acceptance the ordinary understanding bears some nontrivial evidential relation to the object of philosophical interest, and that the role of moral judgment is a genuine part of the ordinary understanding – not simply the result of some bias.

are descriptive views of happiness, which hold that being happy is simply a matter of experiencing the right sorts of psychological states. We have argued that the way we ordinarily understand happiness deviates in important ways from both of these views.

The ordinary understanding of happiness turns out to be a surprisingly sophisticated combination of feeling good and *being* good. Perhaps one reason that these two opposing views have persisted in philosophical discussion is that both views captured different components of the way we ordinarily understand happiness.

References

Annas, J. (1993). *The Morality of Happiness* New York: Oxford University Press.

Annas, J. (2011). *Intelligent Virtue* New York: Oxford University Press.

Beebe, J.R. & Buckwalter, W. (2010). 'The Epistemic Side-Effect Effect.' *Mind and Language* 25(4): 474-498.

Chituc, V. (2012). 'Emotions as Dual Character Concepts.' Unpublished Manuscript, Yale University.

Cooper, J. (1987). 'Contemplation and Happiness: A Reconsideration' *Synthese* 72(2): 187-216.

Crisp, R. (2008). 'Well-Being.' *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2008/entries/well-being/>.

Cushman, F.A., Knobe, J. & Sinnott-Armstrong, W. (2008). 'Moral Appraisals Impact Doing/Allowing Judgments.' *Cognition* 108(1): 281-289.

Deci, E. L., & Ryan, R. M. (2012). 'Self-determination Theory.' In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of Theories of Social Psychology: Vol. 1* . Thousand Oaks, CA: Sage.

Diener, E., Scollon, C.N., & Lucas, R.E. (2003). 'The Evolving Concept of Subject Well-being: The Multifaceted Nature of Happiness. *Advances in Cell Aging and Gerontology, 15*: 187-219.

Dybikowski, J. C. (1981). 'Is Aristotelian Eudaimonia Happiness?' *Dialogue* 20(2): 185-200.

Feldman, F. (2010). *What Is This Thing Called Happiness?* New York: Oxford University Press.

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

Gilbert, D. T. (2006). *Stumbling on Happiness.* New York: Alfred A. Knopf.

Haybron, D. M. (2000). Two Philosophical Problems in the Study of Happiness. *The Journal of Happiness Studies*, 1(2): 207–225.

Haybron, D. M. (2008). *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being.* New York: Oxford University Press.

Haybron, D. M. (2010). "Mood Propensity as a Constituent of Happiness: A Rejoinder to Hill," *The Journal of Happiness Studies*, 11: 19–31.

Haybron, D. (2011). 'Happiness' *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2011/entries/happiness/>.

Hitchcock, C. & Knobe, J. (2009). 'Cause and Norm.' *Journal of Philosophy* 11: 587-612.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kashdan, T., Biswas-Diener, R., & King, L. (2008). 'Reconsidering Happiness: The Costs of the Eudemonic-Hedonic Distinction.' *Journal of Positive Psychology* 3: 219-233.

Kekes, J. (1982). Happiness. *Mind*, 91: 358–376.

Knobe, J. (2003). 'Intentional Action and Side Effects in Ordinary Language.' *Analysis* 63: 190-193.

Knobe, J. & Roedder, E. (2009). 'The Ordinary Concept of Valuing.' *Philosophical Issues*19:1, 131-147.

Knobe, J., Prasada, S. & Newman, G. (forthcoming). 'Dual Character Concepts and the Normative Dimension of Conceptual Representation.' *Cognition*.

Kraut, R. (1979). 'Two Conceptions of Happiness.' *The Philosophical Review* 88(2): 167-197.

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.

Lyubomirsky, S., King, L.A., & Diener, E. (2005). 'The Benefits of Frequent Positive Affect: Does Happiness Lead to Success?' *Psychological Bulletin*, 131: 803-855.

May. J. & Holton, R. (2012). 'What in the World is Weakness of Will?' *Philosophical Studies* 157(3): 341–360.

McCloskey, M., & Kohl, D. (1983). 'Naive Physics: The curvilinear impetus principle and its role in interactions with moving objects.' *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9(1): 146-156.

McCloskey, M., Washburn, A., & Felch, L. (1983). 'Intuitive Physics: The straight-down belief and its origin.' *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9(4): 636–649.

Nadelhoffer, T. (2004). 'On Praise, Side Effects, and Folk Ascriptions of Intentionality.' *Journal of Theoretical and Philosophical Psychology* 24: 196-213.

Pettit, D. & Knobe, J. (2009). 'The Pervasive Impact of Moral Judgment.' *Mind & Language* 24(5): 586-604.

Phillips, J. & Knobe, J. (2009). 'Moral Judgments and Intuitions about Freedom.' *Psychological Inquiry* 20: 30-36.

Phillips, J., Misenheimer, L., Knobe, J. (2011). 'The Ordinary Concept of Happiness (and others like it).' *Emotion Review* 3(3): 1-3.

Roxborough, C. & Cumby, J. (2009). 'Folk Psychological Concepts: Causation.' *Philosophical Psychology* 22(2): 205-213.

Ryan, R. M., & Deci, E. L. (2000). 'Self-determination Theory and The Facilitation of Intrinsic Motivation, Social Development, and Well-being.' *American Psychologist* 55: 68-78.

Ryff, C. D., & Singer, B. (1998). 'The Role of Purpose in Life and Personal Growth in Positive Human Health.' In P.T.P. Wong &. P. S. Fry (Eds.), *The Human Quest for Meaning: A Handbook of Psychological Research and Clinical Applications*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Seligman M. & Csikszentmihalyi M. (2000). 'Positive Psychology: An Introduction.' *American Psychologist* 55: 5-14.

Sidgwick, H. (1907/1966). *The Methods of Ethics*. New York: Dover Publications.

Suikkanen, J. (2011). An improved whole life satisfaction theory of happiness. International Journal of Wellbeing, 1(1), 149-166.

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

Sumner, L. W. (2000). 'Something In Between.' In R. Crisp and B. Hooker (eds.) *Well-Being and Morality (*pp. 1-19). New York: Oxford University Press.

Tiberius, V. (2008). *The Reflective Life*. New York: Oxford University Press.

Tiberius, V. & Plakias, A. (2010). 'Well-Being.' In J. Doris (ed.) *The Moral Psychology Handbook*. New York: Oxford University Press.

Vonasch, A. J., & Baumeister, R. F. (2012). 'True Happiness Is More Than Feeling Happy and Satisfied: People Think "Bad" People Aren't Truly Happy.' Presented at the annual conference for the Society of Personality and Social Psychology, San Diego, CA.

Young, L. & Phillips, J. (2011). 'The Paradox of Moral Focus.' *Cognition* 119: 166-178.