Ordinary Ethical Reasoning and the Ideal of 'Being Yourself'

Joshua Knobe
Princeton University

Forthcoming in *Philosophical Psychology*

Abstract: The psychological study of ethical reasoning tends to concentrate on few specific issues, with the bulk of the research going to the study of people's attitudes toward moral rules or the welfare of others. But people's ethical reasoning is also shaped by a wide range of other concerns. Here I focus on the importance that people attach to the ideal of *being yourself*. It is shown that certain experimental results —results that seemed anomalous and inexplicable to researchers who focussed on moral rules and concern for the welfare of others — can be explained quite elegantly as the product of people's attachment to the ideal of 'being yourself.' The success of this explanation then points to the need for a more general inquiry into the role that the ideal of 'being yourself' plays in people's ethical reasoning.

Ordinary Ethical Reasoning and
the Ideal of 'Being Yourself'

The psychological study of ethical reasoning tends to focus almost entirely on a few specific issues. There has been a great deal of research about people's understanding of moral rules (e.g., Darley & Shultz, 1990; Turiel 1983) and a lot about the importance people attach to the welfare of others (e.g., Gilligan 1982; Weiner 1995). Many other issues have been more or less neglected.

This neglect is surprising given the obvious importance of various other kinds of ethical values in contemporary culture. Consider the dialogue that might result if we brought a typical businessman together with the lead singer of a punk band. Surely, the two would disagree on many important questions about right and wrong, duty and obligation, the nature of the good life. But very few of these disagreements could be adequately characterized as debates about moral rules and the welfare of others. The bulk of the disagreement would concern values of a very different sort.

Here we will be focussing on one of these other values: the importance people attach to the notion of *being yourself*. This is the notion that people exploit when they say that it is important to ‘be true to yourself,’ to ‘be who you really are,’ to be ‘real’ or ‘genuine.’ Clearly, this notion plays an extremely important role in popular culture, where it figures prominently in movies, in self-help books, even in advertisements.

And yet, despite the key role that this notion seems to play in people’s lives, it has received relatively little attention from psychologists. Thus, very little is known about the ways in which this notion figures in people's reasoning about how one ought to live. Nor do we have very much in the way of empirical evidence about individual differences in

people's conceptions of what it means to 'be yourself' or about the development of these conceptions over the course of the lifespan.

Of course, even if psychological research has thus far neglected an important aspect of people's ethical reasoning, it might turn out that psychologists have reached a correct understanding of the specific type of ethical reasoning that takes place in the experiments they have been conducting. In other words, it is at least conceivable that there is something about the design of these experiments that causes people to answer the researchers' questions without making any use of the notion of 'being yourself,' hence that the results of these experiments can be correctly interpreted even in the absence of an accurate understanding of the way that people's notion of 'being yourself' influences their ethical reasoning. On the face of it, however, this outcome seems highly unlikely. If the notion of 'being yourself' does indeed play a key role in people's ethical reasoning, it seems extremely probable that people's responses in at least *some* ethical reasoning experiments are best explained in terms of the importance that people attach to this little-understood notion.

Here I focus on one such case. I argue that certain experimental results —results that seemed anomalous and inexplicable to researchers who focussed on moral rules and concern for the welfare of others — can be explained quite elegantly as the product of people's attachment to the ideal of 'being yourself.' The success of this explanation then points to the need for a more general inquiry into the role that the ideal of 'being yourself' plays in people's ethical reasoning.

Background information

Before we go on to discuss the results of specific studies, it will be necessary to give a brief description of the basic framework within which the researchers were working. However, we should note at the outset that the issue we will be investigating has little or no relation to the framework within which the research was originally conducted. Thus, even if the framework as a whole is revealed to be fundamentally mistaken, the particular issue that we are investigating will lose none of its importance.

Actually, we will be concerned with an issue that entered the world of psychological research more or less by accident, when a group of psychologists (who were really trying to study something else entirely) noticed a small yet recurring irregularity in their data and felt compelled to give some explanation of it. The whole phenomenon is widely seen as the product of a “methodological error” or as a mere “technical issue.” But this is to underestimate its significance. The phenomenon in question is really quite important, more important perhaps than the issues at the heart of the framework in which the researchers were working.

One should note, first of all, that these researchers were not engaged in an attempt to formulate theories about the full range of people’s ethical reasoning. Their research was largely confined to one particular aspect of ethical reasoning, and I shall therefore say that they were engaged in a study of people's *morality* or of their *moral reasoning*. Nothing of any great theoretical importance hangs on the definition of these terms. The idea is simply to reserve the phrase 'moral reasoning' for the particular aspect of ethical reasoning — whatever it happens to be — that research conducted within a particular framework was designed to capture.

This framework has its origins in the influential studies of Jean Piaget (1997/1932) on the moral judgments made by children. Piaget found systematic differences between the judgments made by younger children and those made by older children. It appeared, in other words, that the children were following a definite progression, beginning with certain views early in their lives and then developing other, very different views as they got a little older.

Lawrence Kohlberg drew on these early studies to construct his own highly ambitious research program. He argued that moral development could continue throughout a person's lifespan, so that, e.g., a 30 year-old might still be progressing toward more 'mature' forms of moral reasoning. On Kohlberg's theory, moral development consists in a movement through three basic levels, each of which can be divided into two stages (Kohlberg, 1984):

> *Preconventional* (stages 1 and 2): Moral behavior is seen as an effective means to attain ends that are fundamentally nonmoral (e.g., avoiding punishments and obtaining rewards).
>
> *Conventional* (stages 3 and 4): Morality is identified with social convention. Moral behavior stems from an effort to gain social acceptance or from a more abstract loyalty to societal norms.
>
> *Postconventional* (stages 5 and 6): Moral decisions are based on general principles. The individual believes that these principles should be obeyed even when they come into conflict with societal norms.

It is Kohlberg's view that these levels form a universal, invariant sequence, such that no individual can progress to an advanced stage of moral development without first passing

through each of the lower stages and no individual can move from one of the higher stages down to one of the lower ones.

Regression

In 1969, Kohlberg and Kramer published an in-depth study of moral development in the college years. Although their conclusions were generally in keeping with Kohlberg's earlier theory, they did find one surprising trend: many college students appeared to be *regressing*, moving from conventional to preconventional moral reasoning. This phenomenon was termed the "Raskolnikov regression" after the character in Dostoevsky.

Kohlberg and Kramer explained their results in terms of Erikson's (1950) theory of psychosocial moratoria. College students, they claimed, do not face the pressures of parental control. Yet, at the same time, they are sheltered from the responsibilities of adult life. This brief window of freedom leads to a regression into the hedonism of early childhood (Kohlberg & Kramer, 1969).

The regression hypothesis did not last long. Its most obvious drawback was the curious theoretical status of a 'backward' movement in the stage sequence. Kohlberg's original theory posited a monotonic progression through developmental levels. It had been assumed that movement was always forward, never backward — and certainly not as far back as the "Raskolnikovs" (as they came to be called) seemed to be moving.

A second drawback, however, proved even more devastating. The moral reasoning of "regressed" college students showed marked dissimilarities from the reasoning of traditional preconventional subjects, and it seemed clear that the

Raskolnikovs did not fit neatly into the preconventional category. A preconventional subject sees morality as a means by which to gain rewards and avoid punishment. So, for example, one preconventional subject argues that it is not important to save another individual's life unless that individual commands a high social status:

> If someone important is in a plane and is allergic to heights and the stewardess won't give him medicine because she's only got enough for one and she's got a sick friend in back, they should put the stewardess in a lady's jail because she didn't help the important one (Kohlberg, 1984: 190).

Raskolnikovs, like preconventional subjects, do not believe in universally binding principles. They reject the view that the individual must fulfill certain categorical moral duties. But it would be wrong to say that they have no ethical values at all. Consider the following three quotations from subjects that have been labeled "instrumental hedonists":

> 1) I don't think anybody should be swayed by the dictates of society. It's probably very much up to the individual all the time, and there's no general principle except when the views of society seem to conflict with your own views … say forget it and I'll do what I want. (Kohlberg & Gilligan, 1971: 1074)
>
> 2) I am an individualist. I believe in individuals. And no one tells me what is right and what is wrong. People can guide me…, but I have to make my own choice. (Turiel, 1974: 20)
>
> 3) I have a hard time thinking of anything that would be wrong, except … hypocrisy. I try not to be a hypocrite. (Kohlberg, 1984: 448)

These subjects have clearly rejected the search for universally binding moral principles. Yet, while the Raskolnikovs deny the validity of universal moral principles, they cannot be dismissed as "hedonists" or "relativists." In each of the quotations above, there is a

code of values at work — a code that idealizes authenticity and choice, while denigrating conformity and blind obedience to societal norms.

The most plausible account of their responses, it seems to me, is that they are deeply concerned with an ideal that has nothing to do with either conventional societal norms or universal moral principles. More specifically, it seems that they are deeply concerned with the ideal that we have called *being yourself*. In saying this, of course, we are not doing much to clarify precisely what it is that these subjects have come to value, nor are we providing any kind of hypothesis about the factors that led them to value it. However, we are suggesting that their responses were influenced by an ideal that lies outside the realm of considerations that Kohlberg and his colleagues originally set out to study.[1]

Understanding the Ideal

In fact, the ideal of 'being yourself' appears to differ in fundamental ways from the kinds of values that one finds in the Kohlbergian framework. One cannot accommodate this ideal by simply adding it to some list of values that already appear in Kohlberg's theory. It is a different *kind* of value, and we need to approach it using a quite different framework.

The psychological study of morality has traditionally been concerned with a certain class of *reasons*. Thus, suppose an agent is trying to decide between two options, *x* and *y*. One person might say: 'He ought to choose option *x* because it will give him more pleasure.' And another: 'He ought to choose option *y* because it would be more

fair.' Although both people have offered reasons to choose particular options, it seems that only the second is offering a distinctively moral sort of reason.

Experiments in the Kohlbergian tradition are designed to get at people's understanding of these reasons. Subjects are given stories about a hypothetical agent and asked questions about which option this agent should choose. They are then asked to provide detailed justifications for their answers. The aim is to see which reasons people provide, how they balance these reasons against each other, and so on.

Unfortunately, this approach is not well-suited to the study of people's ideal of 'being yourself.' The trouble is that the ideal of 'being yourself' cannot be associated with a privileged class of reasons. One person might attain this ideal by seeking pleasure, another by trying to do what is fair. More importantly, there is no guarantee that two agents will equally attain the ideal even if both agents act for exactly the same reasons. Consider two poets both of whom reject a certain verse on the grounds that it is too sentimental. It could be that one of the poets is thereby 'being himself' while the other is not. The key question here is not which specific reason the poet acted on but what relation that reason has to the poet's own self.

This issue comes out clearly in the three quotations above. There, the experimenters are asking subjects to say which option an agent should choose and what reasons he has for choosing it. But the subjects refuse to specify a privileged class of reasons. Instead, they respond by talking about the relation in which the agent's reasons should stand to his or her self and the process by which the agent should use those reasons to arrive at a decision.

In particular, subjects seem to be drawing a distinction between cases in which we truly *choose* to perform an action and cases in which we perform an action because (as one subject puts it) we have been 'swayed by the dictates of society.' Perhaps one can get a better understanding of this distinction by applying it to our case of the two poets. We might suppose that both poets reject a verse on the grounds that it is too sentimental but that each poet arrived at this reason in a different way. The first poet has thought deeply about the issue and, after many years of anguished deliberation, decided to banish all traces of sentimentalism from his work. The second has never really even considered the issue — his teachers told him that sentimentalism should be avoided, and he just assumed that they were right. Although both poets perform the same action for the same reason, it might be felt that only the first is truly being himself.

Here one might say: 'The problem with the second poet is that there is a sense in which it was not really *he* who chose to avoid sentimentalism. His decision to reject the verse does not reflect anything fundamental about the kind of person he is. It is simply a reflection of the fact that he happened to take certain classes from certain teachers. To truly "be himself," he would have to think more deeply about the issue and make his own choice between the competing aesthetic approaches.'[2]

Of course, this simple characterization of the ideal of 'being yourself' will not be enough to satisfy most philosophers. One wants to know a great deal more about what kinds of choices are relevant here, how a person can pick up beliefs from others in a way that renders her less truly 'herself,' and so forth. But these are questions for another paper. Our aim here is simply to show that many of the subjects labelled 'Raskolnikovs'

are, in fact, people who insist that the agent in hypothetical dilemmas try to be him- or herself.

Initial support for this conclusion came from the work of Kohlberg and Kramer (1969), where subjects who had been classified as 'instrumental hedonists' seemed actually to be giving voice to a distinctive ideal. Further research by Keniston (1970) and Turiel (1974) both confirmed this finding. Each described the Raskolnikovs as enemies of hypocrisy and champions of individualism. Yet neither sees the Raskolnikovs as advocates of a distinctive ethical view. In fact, both contend that the Raskolnikovs are in the midst of a short-lived phase and will soon regain their belief in binding moral principles. For Keniston, the Raskolnikovs' beliefs are a "temporary regression to moral cynicism" (Keniston, 1970). For Turiel, by contrast, they are a transitional stage between conventional and postconventional reasoning.

Transition

Turiel argues that Raskolnikovs are caught between conventional and postconventional modes of thought. Finding no rational basis for prevailing societal norms, they have rejected conventional moral assumptions and have begun the search for non-relative moral principles. Although they have not yet discovered such principles, they soon will, and at that point, they will become full-fledged postconventional moral thinkers. As it is, they are "confused," and their confusion breeds "disequilibrium, … conflict, contradiction and inconsistency." (Turiel, 1974) This approach was eventually adopted by Kohlberg himself, who began classifying Raskolnikovs as "stage $4^1/_2$" subjects, halfway between conventional and postconventional thought (Kohlberg, 1984).

Why has the Raskolnikov position been classified as a “transition” and not as an independent “stage”? In developmental theory, these two terms have fairly definite meanings (Kohlberg, 1971; Perry, 1968). A developmental *stage* is supposed to be a relatively stable state, the sort of state in which a person might remain for a long period of time. A *transition* is a relatively unstable position between two stages. A person can be considered ‘in transition’ if he or she has recently left one stage and can be expected to enter the next stage relatively soon.

So if Raskolnikovs are, in fact, in the midst of a transition between conventional and postconventional moral thought, it seems that two conditions should apply (Fishkin, 1984):

1. Raskolnikovs should adopt post-conventional thinking within a reasonable period of time.

2. Before becoming Raskolnikovs, they should have reasoned at the conventional level.

There is little evidence that the transition hypothesis meets either of these conditions.

At first, it was assumed that Raskolnikovs were advocates of an unstable position and that they would swiftly find their way back to the easily-classifiable stages of Kohlberg’s system (Kohlberg & Kramer, 1969), but by 1971 Kohlberg was forced to retract this view and conclude that Raskolnikovs might indeed be quite stable in their moral orientation (Kohlberg & Gilligan, 1971). Still, he continued to maintain that Raskolnikovs were involved in a transition between moral stages. This conclusion seems to imply a particularly problematic classification of the great thinkers who would probably be classified as Raskolnikovs under Kohlberg’s system — thinkers like

Nietzsche, Kierkegaard and Sartre. Shall we say that such thinkers spent large portions of their adult lives in a "transitional state"?

The second condition has been even more problematic. It seems clear that Raskolnikovs could not be in the midst of a transition to postconventional thought if they were already engaging in postconventional thought before they became Raskolnikovs. Yet Kohlberg's studies revealed numerous subjects who appeared to be doing just that. In fact, he found that before entering "stage $4^1/_2$," Raskolnikovs tended to be the *most* morally mature of his subjects (Kohlberg, 1984).

Kohlberg's first reaction was to argue that their must have been some sort of methodological error and that the subjects "which we had scored stage 5 were not really principled or stage 5" (Kohlberg, 1973, p. 36). Ultimately, however, he admitted that subjects had, in fact, become Raskolnikovs after attaining stage 5 moral reasoning. He claimed to have "given up calling subjectivism [Kohlberg's word for the Raskolnikov morality] a transitional stage" (Kohlberg, 1984). Yet he still maintained his conviction that Raskolnikov moral reasoning could be analyzed in terms of his system of stages.

A Portrait

Looking through the various attempts that have been made to explain the Raskolnikovs' moral reasoning, one notices an important commonality: almost all researchers attributed their unusual responses to some sort of *problem*. Be it "regression," "confusion" or "disequilibrium," it is always some underlying psychological flaw that is supposed to lead people to adopt the Raskolnikov position.

What we find here is a highly unusual approach to the study of moral reasoning. Normally, if there is a dispute between advocates of two different forms of moral reasoning, it would be odd for one side to accuse the other of having psychological difficulties. Thus, many deontologists think that utilitarianism is a horribly misguided position, but they don't go on to claim that utilitarianism must be the product of acute childhood trauma. Even if utilitarianism is regarded as a big mistake (as it sometimes is), it is nonetheless treated as an intellectual position and not as the symptom of a disease. Not so for the Raskolnikovs. Their views are treated as some kind of psychological problem — the kind of thing that might perhaps be "cured" by a trained therapist.

Still, we should not be too quick to dismiss the prevailing approach to the Raskolnikov position. It is possible, after all, that the Raskolnikovs really are suffering from some psychological problem and that this problem is somehow at the source of their unusual views. So let us turn to the data.

In 1968, Haan, Smith and Block published a study of the moral reasoning of college students. Their main purpose was to examine the relationship between moral reasoning and involvement in political protests. As predicted, they found that postconventional students were more likely to engage in political protests than their conventional peers, but the study also unearthed a more surprising result: Raskolnikovs have an especially high rate of political activism.

To be fair, the study did find that Raskolnikovs were less politically active than postconventional students. But it also found that Raskolnikovs were more likely to act as fundraisers for charitable organizations and equally likely to work as tutors. In fact, while

postconventional students worked in more organizations, Raskolnikovs showed more intense involvement in the organizations of which they were members.

The study also found that mothers of Raskolnikovs were significantly more educated than mothers of students in any other category. And student reports indicated that the mothers of Raskolnikovs were the most likely to expect a lot from their children and the least likely to make fun of them. The students were asked to rate on a 7-point scale the degree to which their mothers "felt I was a disappointment." The average rating given by Stage 3 (conventional) students was 2.1. For postconventional students, the average was 3.9. Among Raskolnikovs, the average rating was a mere 1.8.

As one might expect, the study showed that Raskolnikovs attribute little importance to altruistic behavior. But they appear to be anything but valueless nihilists. Like the postconventional subjects, they have a definite vision of what it means to be a good person. The difference lies not in the *degree* to which they aspire to an ideal but rather in the *content* of that ideal. Raskolnikovs described the ideal individual as playful, free, artistic, doubting, stubborn, uncompromising and aloof. These traits play no role in Kohlberg's scheme, but for Raskolnikovs, they are the hallmarks of a good life.

All in all, the study paints a very flattering portrait of its Raskolnikov subjects. They have loving families, genuine political involvements and strong individualistic ideals. Yet the researchers insist on categorizing Raskolnikovs as "instrumental hedonists." In fact, they view Raskolnikovs as victims of a "moral regression" that stems from "affective disruption of social units" or "personal crises." Needless to say, the authors provide no evidence for these conclusions.

Meta-Ethics

Here I should mention one of the few researchers who takes the Raskolnikovs seriously, treating them as advocates of a distinctive position about the nature of the good life.

Fishkin (1984) seeks to clarify the issue by applying the traditional philosophical distinction between *normative ethics* and *meta-ethics*. The field of normative ethics is concerned with questions about what is good and bad, how people should behave, etc., etc. Meta-ethics is then concerned with the ultimate grounds and the logical status of the claims made in normative ethics. Fishkin argues that these two fields are essentially independent. If we wanted to know about a person's normative ethics, we might ask a question like: "Is it more important to help one's own friends or to obey the law?" Then, no matter what response the person gave, we could follow up with a meta-ethical question like: "Is that response just an expression of your own cultural traditions, or does it reflect an objective truth?"

So in Fishkin's view, it would be best to assess each person's moral reasoning along two separate dimensions: reaching an understanding of his or her normative ethics and then, independently, of his or her meta-ethics. He goes on to argue that the views of the Raskolnikovs are, in essence, meta-ethical. The Raskolnikovs reject the idea that morality can be grounded in anything 'universal' or 'objective,' believing instead that morality rests ultimately on each individual's own thoughts and feelings. On Fishkin's theory, the Raskolnikovs' position is really compatible with any stage in the Kohlbergian categorization scheme, since that scheme is concerned primarily with issues in normative ethics. One Raskolnikov might say that morality consists in obeying certain rules, another

that morality is a matter of helping one's family and friends. What makes them both Raskolnikovs is that, when you ask them for the ultimate grounds of the opinions they offer, they both say something like: "It all comes down to the feelings of the individual, and that's just how I feel."

Fishkin is drawing a helpful distinction here, but I think that the Raskolnikov position cannot be confined to the realm of meta-ethics. After all, there are two very different contexts in which a person might say something like "It's all very much up to the individual." Suppose we ask whether a given person — call him Heinz — should choose option *x* or option *y*. It is possible, as Fishkin suggests, that some subjects would say, "Heinz should choose option *x*" and then, when asked about the meta-ethical status of this claim, would answer, "It's very much up to the individual, and that's just how I feel." But there is also another possibility to be considered — this one at the level of normative ethics. The subject might suggest neither *x* nor *y* but simply say: "It's all very much up to the individual. Heinz should act on his own thoughts and feelings." This latter claim has nothing to do with meta-ethics. It is a straightforward normative claim. The idea is that Heinz ought to act on his own thoughts and feelings, and if he doesn't — if he somehow fails to act on his thoughts and feelings — he is doing something wrong. One could then ask about the meta-ethical status of this judgment, but that would be a separate question, and there is no reason to suppose that it would always elicit the same answer. Some subjects might regard their judgment as the reflection of an objective truth, some as a personal preference, and some — probably quite a few — just won't have any views at all about meta-ethics. Even if they believe very strongly in a certain vision of the ideal

life, they might never have given much thought to questions about the epistemological or metaphysical status of their beliefs.

We have now distinguished two different kinds of responses that might lead a subject to be classified as a Raskolnikov. First, there is the meta-ethical claim that the ultimate grounds of moral judgment lie in the thoughts and feelings of the individual. Second, there is the normative ethical claim that each individual ought to act on his or her own thoughts and feelings. In the Kohlbergian scheme, these two claims get lumped together and treated as different strands of a single basic approach to moral reasoning. But the two should be separated, and in what follows, I will be concerned exclusively with the second — the normative claim.

Interim Conclusion

Thus far, we have discussed three explanations of the Raskolnikovs' position — 'regression,' 'transition' and 'meta-ethics.' All three proceed by applying a conceptual framework that was originally developed as part of an effort to understand people's reasoning about moral rules or the welfare of others.

So, for example, in our effort to understand people's concern for moral rules and the welfare of others, it may be helpful to develop a framework according to which people distinguish between hedonistic desires and moral obligations. (Indeed, this strikes me as quite a plausible view about that particular issue.) The regression hypothesis then takes this framework and uses it to explain Raskolnikov subjects. In particular, the hypothesis claims (rather implausibly) that Raskolnikovs think people ought to follow their hedonistic desires at the expense of their moral obligations.

Similarly, when we are engaged in a study of people's concern for moral rules and the welfare of others, it may be helpful to develop a framework that draws distinctions among a certain number of common meta-ethical viewpoints. The meta-ethics hypothesis is an attempt to take this very framework and use it to explain the Raskolnikovs' behavior. Although the hypothesis may successfully explain the behavior of certain subjects who have been classified as 'Raskolnikovs,' it fails to explain the views of those subjects who appear to be advancing a definite normative claim.

To understand the views of such subjects, it will not be enough merely to borrow concepts and distinctions from existing theories of moral reasoning. What we need is a theory that describes and explains people's concern with the ideal of 'being yourself,' and such a theory will presumably involve concepts and distinctions that play no role at all in existing theories.

But, of course, all this is to be expected. As we noted above, people's ethical reasoning is shaped by a wide variety of different concerns. People are certainly concerned with moral rules and with the welfare of others, but their ethical reasoning is also influenced in a very fundamental way by other, very different concerns — concerns that have little or no relation either to moral rules or to the welfare of others. It therefore seems unlikely that any single conceptual framework could serve, all by itself, to explain the entirety of people's ethical reasoning. Rather, we should expect to find that adequate theories of people's ethical reasoning must make use of a quite heterogeneous variety of different concepts — many of them outside the realm of what has been studied by existing research programs.

On Experimental Design

The conclusion we have been defending thus far is often obscured by certain peculiarities of the design of ethical reasoning experiments. Subjects in such experiments are never asked very general questions like “How shall we live?” Instead, they are presented with specific moral dilemmas and asked to provide solutions. The data obtained in the experiments therefore depend in a crucial way on the particular dilemmas that the researchers choose to pose.

Strangely enough, the dilemmas used by researchers in the Kohlbergian tradition are not drawn from anything in contemporary culture. Many of them are modified versions of dilemmas that appeared in medieval casuistry, and they therefore thematize one of the central problems of medieval ethics: the potential conflict between universal moral principles and the laws of a given nation. So, for example, one question concerns a man whose wife needs a certain medicine to stay alive. The local druggist is charging some exorbitant price, and he simply cannot afford to buy it. Should he break in and steal the medicine?

With questions like these, it’s a wonder that subjects ever even mention issues like sincerity or personal authenticity. The whole experimental design seems to force the subject into a certain type of thinking. One is quite clearly supposed to weigh the sanctity of life against the rule of law. To the extent that one starts talking about something else entirely, one seems somehow to be violating the implicit rules of the experiment.

How then shall we understand those few subjects who, even in this artificial situation, based their whole response on concepts like “being yourself,” “making your own choices,” or “avoiding hypocrisy”? One option would be to suppose that they

represent some kind of alienated fringe group, obsessed with issues that most people find totally unimportant. But there is another, more plausible explanation. Maybe a great portion of the American population shares the concerns of these so-called "Raskolnikovs," and it is only the specifics of the experimental design that keep subjects from mentioning such issues in response to the dilemmas.

By manipulating the specifics of the experimental situation, psychologists can have a great impact on the concepts people use in their answers. I might be able to construct a dilemma that many people addressed in terms of concepts like "sticking up for yourself," "courage," and "being a real man." Or I could create a dilemma that led people to use concepts like "humility," "sin," and "God." Now if I really went out of my way to create a situation in which it seemed totally inappropriate to mention God, I might be able to get mentions of God down to, say, 1% of all responses. But this would not show that the other 99% of the people thought God was irrelevant to ethical reasoning. For every person who mentioned God in the artificial situation I created, there might be 40 who would have mentioned God if I had changed the situation in some subtle respect.

Perhaps an analogous effect applies to people's use of the ideal of 'being yourself.' When we observe that very few people appeal to this ideal in a particular experimental situation, we should not immediately conclude that it is some idiosyncratic value to which most people attach no importance. Perhaps most contemporary Americans attach at least some importance to the ideal of 'being yourself' (treating it as one value among many), and the so-called 'Raskolnikovs' differ only in that they apply this ideal in a broader range of situations.

If we want to know which concepts are most essential to people's reasoning, it is not enough to construct hypothetical dilemmas and then present them to subjects. It is also necessary to figure out which dilemmas people are actually concerned with. For example, it might be helpful to take a look at popular entertainment. Just watch a few modern movies, some after-school specials on TV, maybe a best-selling novel. You will undoubtedly encounter a dilemma that looks more this:

> The big dance contest is coming up, and Jake thinks he has a good chance of winning. But there is a problem. The judges expect contestants to dance in the measured, traditional style that has been practiced for centuries. Jake hates that way of dancing. He has been perfecting a new style — all his own — that is more expressive and fiery, even a bit sexually provocative. Jake's parents and friends try to dissuade him from using his new style at the contest. "It's pointless," they say. "You'll be sacrificing your chances for a successful dance career, and you'll disgrace all of us with this foolishness." Only Jake's girlfriend believes that his style is worth pursuing. Just before the contest begins, when Jake has already decided to follow the advice of his parents and friends, she gives him an inspiring speech about how he has to "hold onto the dream."
>
> Jake mounts the stage and stands before the judges. What should he do?

Faced with a dilemma like this, the audience would be far more likely to think in terms of concepts like *being yourself.* And it behooves us, as psychologists and philosophers, to make some attempt to understand these concepts and the roles they play in people's lives.

Notes:

[1] Note that we can remain strictly neutral about the claim, made by Gilligan (1982) and Haidt (2001) among others, that the Kohlbergian theory is wrong even about the very phenomena it was designed to explain. Our aim is simply to point to a class of phenomena that lie outside the scope of the Kohlbergian theory. Thus, our chief claims are entirely orthogonal to the debate between Kohlberg and his more radical critics.

[2] With its emphasis on personal choice, this sentiment is reminiscent of Sartre's (1956/1943) concept of authenticity. But there are a number of important differences. First, Sartre would deny that we have any antecedently-existing 'self' which guides our choices and is reflected in them. Second, Sartre would say that we are in fact always making our own choices, even when we act most inauthentically — the only question is whether we acknowledge the choices we are making.

Acknowledgements

An abridged version of this paper was presented at a meeting of the Society for Philosophy and Psychology, where Owen Flanagan served as commentator. I am grateful to Bertram Malle and Phillip Zimbardo for helpful comments on an earlier ancestor of the present paper and to Gilbert Harman and three anonymous reviewers at *Philosophical Psychology* for comments on the penultimate draft.

Works Cited

DARLEY, J. M. & SHULTZ, T. R. (1990). Moral rules: Their content and acquisition. *Annual Review of Psychology*. *41*, 525-556.

ERIKSON, E. (1950). *Childhood and society*. New York: Norton.

FISHKIN, J. (1984). *Beyond subjective morality*. New Haven: Yale University Press, 1984.

GILLIGAN, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.

HAAN, N., SMITH, B., & BLOCK, J. (1968). The moral reasoning of young adults. *Journal of Personality and Social Psychology, 10*, 183-201.

HAIDT, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review. 108*, 814-834.

KENISTON, K. (1970). Student activism, moral development and morality. *American Journal of Orthopsychiatry, 40*, 577-92.

KOHLBERG, L. & GILLIGAN, C. (1971). The adolescent as philosopher. *Daedalus, 100*, 1051-1086.

KOHLBERG, L. & KRAMER, R. (1969). Continuities and discontinuities in child and adult moral development. *Human Development, 12*, 93-120.

KOHLBERG, L. (1973). Continuities in childhood and adult moral development revisited. In P. B. BALTES AND K. W. SCHAIE (Eds.) *Life- span developmental psychology: personality and socialization.* New York: Academic Press, pp. 180-204.

PERRY, W. G. (1968). *Forms of intellectual and ethical development in the college years: a scheme*. New York: Holt, Rhinehart and Wilson.

PIAGET, J. (1997). *The moral judgement of the child.* New York: Simon & Schuster. (Originally published in 1932.)

SARTRE, J.-P. (1956). *Being and nothingness.* Trans. Hazel Barnes. New York: Washington Square Press. (Originally published in 1943.)

TURIEL, E. (1974). Conflict and transition in adult moral development. *Child Development, 45*, 14-29.

TURIEL, E. (1983). *The development of social knowledge: Morality and convention.* Cambridge: Cambridge University Press.

WEINER. B. (1995). *Judgments of responsibility*: A foundation for a theory of social conduct. New York: Guilford Press.