

Reading Conflicted Minds: An Empirical Follow-up to Knobe and Roedder

Chad Gonnerman
Indiana University, Bloomington

[*Note*: This is a preprint of an article whose final and definitive form has been published in the *Philosophical Psychology*, Vol. 21, No. 2 (2008): 193-206. Please refer to published version. *Philosophical Psychology* is available at: <http://journalonline.tandf.co.uk/>]

1. Introduction

While the early reviews are mixed – ranging from the criticism that it’s all sizzle, no steak (e.g., Kauppinen, 2007) to the kinder view that it makes for a decent philosophical starter (e.g., Nahmias *et al.*, 2005) – one thing that we can safely say about the offerings of experimental philosophy is that they often surprise. Here is a brief sample of the unexpected. Weinberg *et al.* (2001) found that intuitions in response to Gettier cases can vary across cultural and socioeconomic groups; Nichols and Knobe (forthcoming) report that intuitions concerning the compatibility of determinism and moral responsibility can be influenced by a thought-experiment’s affective content; and Swain *et al.* (forthcoming) revealed that intuitions in a Truetemp case can fluctuate according to which other thought-experiments were first considered. Thanks to Joshua Knobe and Erica Roedder, we now have another item to chew over.

Recently they found that folk attributions of valuing can vary according to the perceived moral goodness of the object (Knobe & Roedder, 2006). For example if the subject judges that racial equality is morally good, the evidence suggests that she will be more inclined to attribute the valuing of it to some other person than if she does not judge

so. Now, that is an interesting find – clearly interesting. But what is not so clear is what it means.

Knobe and Roedder have a controversial answer. It means that MORAL GOODNESS is a conceptual feature of the concept VALUING.¹ But I think that their argument proves too much. My goal is to present an empirically informed argument that supports that thought. A bit more specifically, our own research [*reference suppressed for review*] points to the existence of what we might call “K&R-patterns” in folk attributions of desires and moral beliefs, where more generally a *K&R-pattern* arises whenever there is a pattern in folk attributions of pro-attitudes such that a psychologically conflicted person is more likely to be attributed the pro-attitude toward an object when the object is judged to be morally good than when it is judged otherwise. I will argue that the existence of K&R-patterns in desiring attributions, coupled with the sorts of considerations that motivate Knobe and Roedder, should lead us to conclude that DESIRING has a moralized interior too. But that, as we will see, is probably mistaken. Therefore, it is also probably mistaken to conclude that MORAL GOODNESS is a conceptual feature of VALUING, at least on the basis of the evidence that Knobe and Roedder report. I think that something else is going on in both their findings and ours. At the end I will sketch the very beginnings of what that something else might be.

¹ Exactly how MORAL GOODNESS is supposed to be in the concept VALUING will partially depend on which model is best for understanding the structural relation between concepts and their features. As Laurence and Margolis (1999) point out, there are containment models and inferential models of this relation. If the former is best, then MORAL GOODNESS would be literally part of VALUING. Consequently the tokening of VALUING requires the tokening of MORAL GOODNESS. If, on the other hand, the latter is best, then MORAL GOODNESS would be only inferentially connected to VALUING but this connection need not be activated when either is tokened. Consequently VALUING may be tokened without MORAL GOODNESS being tokened. To help us keep the two relations straight, we might say that on a containment model Knobe and Roedder maintain that MORAL GOODNESS is a *constituent* of VALUING and on the inferential model they maintain that it is a *component* of VALUING. I will use ‘feature’ and ‘element’ as neutral between both models.

2. Knobe and Roedder on VALUING

Before we get to the paper's main findings, we should try to get a clearer understanding of Knobe and Roedder's claim that MORAL GOODNESS is a feature of VALUING, both what motivates it and what it precisely means. Beginning with the former, I discuss some of the relevant experimental findings.

2.1. Knobe and Roedder's Experimentally Driven Argument

Knobe and Roedder remark that many philosophers operate with the assumption that “the concept of valuing can be defined in purely descriptive, non-normative terms” (2006, p. 1). David Lewis (1989), for instance, suggests that we can identify states of valuing something with states of desiring to desire it. And, after taking his readers through a long process of Gricean “creature creation”, Michael Bratman (2000) ends up proposing that there are two kinds of valuing states: those that can be identified with considered desires and those that involve self-governing policies of treating desired objects as ends for action. Both views assume that real states of valuing are ultimately to be identified with other complex psychological states. Perhaps because of this assumption, Knobe and Roedder see them as also assuming that the concept VALUING is a purely descriptive, folk-psychological concept.

But there is a problem, argue Knobe and Roedder. Some empirical evidence suggests that MORAL GOODNESS is a moderately weighted feature of VALUING. In two different studies, they found that when subjects are presented with vignettes describing a deeply conflicted individual – one for whom some of the explicitly attributed

psychological states suggests that she values a certain object and others suggests that she does not – they are significantly more likely to say that the individual values the object when they perceive it to be morally good than when they perceive it differently. If MORAL GOODNESS is a moderately weighted feature of VALUING, these K&R-patterns are what we might expect. To see why, consider their first study involving the racially conflicted Georges.

For those unfamiliar with the originals, here is the first thought-experiment, the case of *racist George*.

George lives in a culture in which most people are extremely racist. He thinks that the basic viewpoint of people in this culture is more or less correct. That is, he believes that he ought to be advancing the interests of people of his own race at the expense of people of other races.

Nonetheless, George sometimes feels a certain pull in the opposite direction. He often finds himself feeling guilty when he harms people of other races. And sometimes he ends up acting on these feelings and doing things that end up fostering racial equality.

George wishes he could change this aspect of himself. He wishes that he could stop feeling the pull of racial equality and just act to advance the interests of his own race.

Does George value racial equality? Notice how conflicted he is. He consciously believes that he ought to act in the interests of his own race, but he has these lingering egalitarian feelings. He wishes he could rid himself of those feelings, but he experiences guilt when he acts like a racist. If MORAL GOODNESS is a moderately weighted feature of VALUING, then its activation might be just enough to tip the scales, barely passing the threshold for a valuing attribution. And the evidence indicates that is what we see. Knobe and Roedder report that their subjects tended to say that George values racial equality.

We find a similar state of conflict in their second thought-experiment, the case of *liberal George*. The major difference is that where we find racist and egalitarian contents in racist George's psychology, we have the reverse here.

George lives in a culture in which most people believe in racial equality. He thinks that the basic viewpoint of people in this culture is more or less correct. That is, he believes that he ought to be advancing the interests of all people equally, regardless of their race.

Nonetheless, George sometimes feels a certain pull in the opposite direction. He often finds himself feeling guilty when he helps people of other races at the expense of his own. And sometimes he ends up acting on these feelings and doing things that end up fostering racial discrimination.

George wishes he could change this aspect of himself. He wishes that he could stop feeling the pull of racial discrimination and just act to advance the interests of all people equally, regardless of their races.

Does this George value racial *discrimination*? Again, the state of conflict is similar to that described in the first case. But here MORAL GOODNESS should remain inactive, since presumably most of us do not think that racial discrimination is morally good. Thus if MORAL GOODNESS is a moderately weighted feature of VALUING, then its silence should keep the valuing attribution below the necessary threshold. Again, that is what we seem to see. Knobe and Roedder report that their subjects tended to deny that liberal George values racial discrimination.

They conclude that VALUING is not a purely descriptive concept. Just as we might conclude that FEATHERED is a feature of BIRD because of our greater tendency to categorize an object as a bird once we notice that it is feathered compared to when we do not, so goes MORAL GOODNESS and VALUING, or so it seems.

2.2. Clarifying the Claim: Two Kinds of Conceptual Investigations

It is important to read their claim that MORAL GOODNESS is feature of VALUING correctly. Many philosophers are likely to misinterpret Knobe and Roedder, taking them to be saying something far stronger than they want.

When philosophers talk about the internal features of a concept, they usually have in mind its semantic features, especially its reference- (or perhaps extension-) determining features. The reason is that most philosophers see themselves as fundamentally concerned with explaining the nature of real things in the world, and reference provides that the connection between concepts and the world. So if a philosopher were to insist that KNOWLEDGE is JUSTIFIED TRUE BELIEF, she probably means to say what it really takes for some bit of the world to count as knowledge. True, in the hopes of securing the reliability of philosophical intuition, these concepts and their features are often understood to be psychologically real and to have some role in generating intuitive categorization judgments (Ramsey, 1998; Goldman & Pust, 1998; Goldman 2007). But that is of secondary interest to most philosophers. Most are not terribly concerned with the precise nature of the processes and representations that people actually use in deciding whether some bit of the world counts as knowledge. What they want to know about is knowledge itself. If philosophical analysis begins with concepts, that involves specifying reference-determining features. That has been the game in the philosophy for some time now, arguably as long as conceptual analysis has been on the scene.

So it is natural for philosophers – we have many years of conditioning to overcome – to read Knobe and Roedder as interested in the referential contours of VALUING, and thus in the truth-conditions of valuing attributions (as, e.g., Kauppinen,

2006, reads them). Of course we cannot read them as doing traditional conceptual analysis. As a bit of experimental philosophy, they have the left armchair, seeking more than the intuitions of the Few, the Proud, the Philosophers. Still, they could be seen as doing semantics in a different form. Borrowing some terminology from Alexander and Weinberg (2007), they would be closer to *intuition populism*, rather than *solipsism* (e.g., Fumerton, 1983; 1995) or *elitism* (e.g., Ludwig, 2007).

But whatever semantic populism has going for it, we should understand that it is not Knobe and Roedder's project (see Roedder & Knobe, 2006; Knobe, 2007a; Knobe, 2007b). In explaining their results, they gloss VALUING as a prototype. Prototypes are in general ill-suited for giving referential values. To use a well-worn example, consider GRANDMOTHER. Prototypical grandmothers have gray hair and, in this part of the world, like to bake cookies. But nobody thinks that being a grandmother depends on hair color and baking preferences. If it were, we would have a lot less grandmothers and a lot more mysterious birthday cards in this world. So we shouldn't read them as saying that MORAL GOODNESS is part of what it metaphysically takes for something to fall under the concept VALUING, and thus that it has something to do with correct uses of 'valuing' (cf. Knobe, 2003).²

They only mean to commit themselves to a claim about our categorization systems. More precisely, they are claiming that when *judging* whether some instance falls under VALUING, we access the MORAL-GOODNESS feature, where our accessing of it is sufficiently robust and universal to be considered genuinely constitutive of VALUING. So

² Because of this, Knobe and Roedder probably shouldn't have said that they are taking on the widespread assumption in philosophy that VALUING is a purely descriptive concept. Philosophers like Lewis and Bratman are probably best interpreted as addressing the metaphysics of valuing.

their claim pertains to our psychology (or mere psychology, as some analytic philosophers might put it before tuning out).³

This is not to say that there are no mysteries remaining with their claim. A major one has to do with the extent to which concepts are involved in categorization (see, e.g., Fodor, 1998; Prinz & Clark, 2004), and thus whether Knobe and Roedder are entitled to draw conclusions about the concept VALUING on the basis of folk categorization judgments. Regardless their claim is clear enough for our purposes. From most any cognitive-science perspective, there are certain psychological structures involved in the generation of categorization judgments. Knobe and Roedder think that there is a certain structure that is reliably and nearly universally accessed when attributions of valuing states are made. Their claim is about that structure and its features, what we might call the *categorizational concept* VALUING. Whether that structure is the same one, or part of a more complex one, that serves other explanatory agendas often given to concepts – for example, whether it plays any role in specifying the truth-conditions of valuing-attributions – is an open question, but not one that we have to resolve here.⁴

³ Of course this is not to say that they should tune out. The nature of the representations and processes that produce intuitions can be relevant to analytic philosophy in many ways. For one, it could give us reason to distrust certain intuitions and their use in philosophical theorizing. To illustrate, suppose that Knobe and Roedder are right in claiming that MORAL GOODNESS is a feature of VALUING and that VALUING is best glossed as a prototype. If so, then certain intuitions are presumably the result of the operations of this prototype. One of Gilbert Harman's (1993, p. 151) counterexamples to David Lewis' (1989) identification of (intrinsic) valuing with (intrinsic) second-order desires may be an example. Perhaps the reason we attribute a state of valuing to the conflicted character in Harman's story, despite her lack of a second-order desire, is that we are talking about an object – listening to Mozart – that has some tinge of moral goodness to it. After all, would we get the same intuition if the object were some guilty pleasure, like listening to the Spice Girls? If so, then there is some reason not to allow that intuition into the foundations of one's philosophical theorizing.

⁴ One theoretical option is to appeal to Laurence and Margolis' hybrid theory, according to which concepts are highly structured mental representations with multiple distinct parts, fulfilling different explanatory agendas (see their 1999; 2003). So, one part of VALUING might encode its defining properties, thus specifying the correct uses of 'valuing', while another part encodes its prototypical properties, thus employed rapid categorization judgments (cf. Osherson & Smith, 1981). On this general account Knobe and Roedder's claim would only be that MORAL GOODNESS is *in* the latter part, VALUING's nonsemantic

With that clarification on hand, we can now ask about the experimental evidence. Have Knobe and Roedder given sufficient evidence for thinking that MORAL GOODNESS is a feature of the categorizational concept VALUING? I think probably not. The following makes for part of my case.

3. The Experimental Evidence: Cases of Conflicted Souls

To address the adequacy of Knobe and Roedder's hypotheses, we conducted two simple surveys. Since we think that something other than a moralized feature on the concept VALUING is responsible for their results, we predicted that we should find similar results, further K&R-patterns, when it comes to other folk-psychological concepts. The study examines whether this is the case with folk attributions of desires, the second with folk attributions of moral beliefs.

3.1. Folk Attributions of Desires

44 students attending undergraduate classes at a large, Midwestern university were recruited. We presented each subject with one of Knobe and Roedder's two vignettes involving the racially conflicted Georges.

Whereas Knobe and Roedder asked about the Georges' states of valuing, we asked about their desires. More specifically our probe was, 'Please indicate the extent to which you agree or disagree with the following sentence: 'Deep down, what George really wants is racial equality [racial discrimination].'' Like Knobe and Roedder,

structure. The danger of such an approach is to keep concepts from becoming monstrously big representational entities while still serving all the explanatory agendas. I am doubtful that this can be done.

subjects' responses were measured using a scale that went from -3 ('definitely disagree') to +3 ('definitely agree'), with 0 identified as 'in between'.

As predicted, in the analysis of the data, we found results similar to Knobe and Roedder's. That is, we found a K&R-pattern in attributions of desires. The mean rating in the first condition was .320 ($N = 25$), and in the second it was -1.05 ($N = 19$). A t -test revealed that the difference between these means scores is statistically significant ($p = .016$). This suggests that in general people are more willing to say that what racist George really wants is racial equality than they are willing to say that what liberal George really wants is racial equality. Can these patterns be found elsewhere?

3.2. Folk Attributions of Moral Beliefs

This second study looked to address that question. For this one, 53 additional undergraduate students were recruited from classes at the same large, Midwestern university. Each received a modified version of one of the vignettes involving our racially conflicted Georges.

In both thought-experiments we eliminated all explicit belief attributions and put in their place statements about what they often say. The rationale for these changes is that we wanted to ask about the Georges' beliefs: basically, does George believe that promoting racial equality [discrimination] is right? The worry is that without these changes subjects could overly rely on these explicit attributions, perhaps through some matching heuristic, in responding to the probe. Here is the complete modified version of the racist-George case:

George lives in a culture in which most people are extremely racist. He often says things that more or less accord with the basic viewpoint of the people in this

culture. That is, he often says that he ought to be advancing the interests of people of his own race at the expense of people of other races.

Nonetheless, George sometimes feels a certain pull in the opposite direction. He often finds himself feeling guilty when he harms people of other races. And sometimes he ends up acting on these feelings and doing things that end up fostering racial equality.

George wishes he could change this aspect of himself. He wishes that he could stop feeling the pull of racial equality and just act to advance the interests of his own race.

After the thought-experiment, each subject received the following probe: Please indicate the extent to which you agree or disagree with the following sentence: ‘Despite what he often says, deep down, what George really believes is that promoting racial equality [racial discrimination] is the right thing to do.’ As in the previous studies, their responses were measured using a seven-point Likert scale.

Again, as predicted, analysis of the data revealed patterns similar to Knobe and Roedder’s. We found that subjects were more likely to say that, deep down, racist George believes that promoting racial equality is the right thing to do than they were willing to say that, deep down, liberal George believes that promoting racial discrimination is the right thing to do. The mean rating in the first condition was 1.192 ($N = 26$). The mean rating in the second condition was $-.286$ ($N = 28$). A t -test revealed that the difference between these means is statistically significant ($p = .001$).

So, like Knobe and Roedder, we found K&R-patterns in our data, in particular, an asymmetry in folk attributions of desires and moral beliefs. These findings are important because they suggest that these patterns can arise elsewhere in folk psychology. And that provides some reason for thinking that something else is generating their results. I argue for that claim in the next section.

4. MORAL GOODNESS is not a Feature of VALUING

If someone wishes to argue that MORAL GOODNESS is a feature of VALUING by appealing to the existence of K&R-patterns, it had better not be the case that similar patterns arise in our attributions of other pro-attitudes. To see this, let's focus on the K&R-patterns in desiring attributions. With asymmetric attributions of both valuing states and desiring states, the pressure is to explain the two similarly, to treat like alike. This leaves us with one of two conclusions: either (1) both concepts have a moralized interior and it is the accessing of the moral feature that explains the empirical results or (2) both lack this moralized interior and it is some extra-conceptual factor that explains the results. The former, or so I will argue, is unattractive when it comes to DESIRING. The problem is that there is no currently popular account of conceptual structure for which it seems terribly plausible to claim that MORAL GOODNESS is a feature of DESIRING.

As mentioned earlier, Knobe and Roedder like to gloss their explanation in terms of prototypes. VALUING, they think, is a prototype and one of its moderately weighted features is MORAL GOODNESS. The usual line on prototypes is that they are mental representations that encode the statistically frequent, usually salient, properties of the category (Rosch, 1978; Laurence & Margolis, 1999). For example, using a listing procedure, one according to which subjects simply write down the features of a category that come to mind when prompted (see Rosch, 1973; Rosch & Mervis, 1975), we may reasonably conclude that the prototypic structure of BIRD includes such usual suspects as FLIES, SINGS, FEATHERED and so on.

The problem for those who wish to take a similar line on MORAL GOODNESS and DESIRING is that it is not clear whether the folk regard states of desiring as typically having the property of moral goodness. In fact, reflection on the sorts of desires that we meet on a daily basis (desires for food, drink, rest, and the bathroom) suggest that many, perhaps most, states of desiring are morally unmarked. They are neither good nor bad. They just cry out for satisfaction. In addition to these, we have the great troublemakers for any attempt to develop a plausible desire-based ethics (e.g., Brandt's (1979) attempt to characterize "rational desires"): morally problematic desires such as those of an alcoholic or a murderer. The point is that with little effort we can come up with many desires which lack the property of moral goodness, which suggests that the folk readily recognize that these two often part ways. Of course I am prepared to be proven wrong here. For example, a listing procedure may reveal that the folk freely associate moral goodness with states of desiring. But I am not willing to bet on it.

So if we want to maintain that MORAL GOODNESS is a feature of DESIRING, I think we will have to recruit a different account of conceptual structure. The obvious candidate here is the theory-theory, since it welcomes the most into conceptual houses. But there is one major problem with this retreat. And we begin to see it once we note, as Knobe and Roedder do (2006, p. 2), that judgments of moral goodness will reveal themselves in the data only under quite special circumstances, namely, when ordinary people confront a seriously conflicted individual, someone like racist George. In our attempt to explain the application of folk-psychological concepts in such special circumstances we should guard against allowing the concerned features into our conceptual houses. Otherwise, we run the risk of making our concepts far too bloated and rendering conceptual features a

theoretically uninteresting notion. After all, it cannot be that every categorization judgment turns on a feature that belongs *in* the deployed concept. To adapt an example from Laurence and Margolis (1999, pp. 72-3), thought-experiments presumably could be devised so that the judgment that this object is smarter than a rock makes all the difference in categorizing it as a bird. But we shouldn't conclude that BEING SMARTER THAN A ROCK is partially constitutive of BIRD, as (say) FEATHERED might be. It is not a sufficiently robust feature. It is better to regard such features, so rarely relied upon, as simply useful information that occasionally helps to issue in a categorization judgment.

What I am getting at, then, is this: maintaining that MORAL GOODNESS is a feature of DESIRING is not terribly plausible. But if it's not plausible here, then it is not there either. Absent any reason for thinking that VALUING is unique, we should treat like cases alike. So we are going to have to look elsewhere to explain the existing patterns.

5. K&R-patterns: The Result of an Egocentric Bias in Mindreading?

The primary problem in explaining the patterns is that there are too many things that one could say as far as these empirical results go. The race to crack K&R-patterns is still early. There are many stories available to the would-be teller. While fully acknowledging that, I want to suggest a way to approach these patterns.

I think progress in coming to understand these patterns can be made if, for the moment, we resist the inclination to tell some story about the precise representational features or processing characteristics that generate these results. In its place, we should ask, Do these patterns look like anything else coming out of psychology? The answer, I think, is yes. It seems to me that these K&R-patterns fit a more general pattern of

psychological results, those that reveal an egocentric, or projection, bias in third-person mindreading. The thought here is that our third-person mental-state attributions are often overly influenced by own (genuine, non-pretend) psychological states (cf. Goldman, 2006, p. 165), a situation likely to arise when we have little information about our target.

Perhaps the existing K&R-patterns are more instances of this bias. To illustrate, consider racist George. When the subject considers the question of what George values, she is probably somewhat puzzled, if only temporarily. The information explicitly contained in the narrative is insufficient for answering the question. Well, so goes the proposal, when in doubt, people tend to rely on their own psychologies, for example, their own valuing states. Presumably our subject values racial equality. Consequently we see that she tends to say that George does too.

Notice that by fitting the existing K&R-patterns within a larger body of results and theory, unlike Knobe and Roedder, we are being quite conservative. Many areas of psychology give us reason to think that we are often egocentric thinkers. Here is just a brief sample. We see this predisposition in the informal reasoning literature on the myside bias, our tendency to gather evidence and evaluate claims or arguments in a self-centered way (Stanovich *et al.*, in press). For instance, it was found that Americans are significantly more likely to claim that the United States should ban a dangerous German car than they are to claim that Germany should ban an equally dangerous American car (Stanovich & West, in press). There is also the psychological literature on projection, our tendency to attribute our characteristics to others (Holmes, 1968). One of the more interesting finds here, given our topic, is that parents tend to project their values on their children (Whitbeck & Gecas, 1988). There is finally the social psychological work on the

“false consensus effect”, our tendency to overestimate our own characteristics when estimating how common they are in a population (Savion, 2008, pp. 78-79). To take a well-known example, it was found that those willing to wear a giant “REPENT” sign estimated that 60% would also feel the same way, while the unwilling guessed that only 27% would feel the same (Ross *et al.*, 1977). (For several other examples, with a special emphasis on mindreading, see Goldman 2006, pp. 165-168.) So there is already good evidence for thinking that egocentric biases are a common part of our cognitive life. We might as well fit the existing K&R-patterns within that body of results and theory, if at all possible.

Before closing, let me briefly comment on a virtue of this proposal. When we step back and attempt to see the existing K&R-patterns as examples of a more general pattern of egocentric thinking in mindreading, the opportunity arises to fit this research within a broader psychological program: the attempt to understand the conditions in which, and the representations and processes responsible for, biased attributions in mindreading. This has two attractive consequences.

First it opens up a rich source of theory to be used in developing candidate explanations. For instance, we might draw on the basics of Nichols and Stich’s (2003) cognitive architecture in attempting to understand the causes of the existing K&R-patterns. In addition to the process of default belief attribution – that of including all our beliefs (or something like that; see p. 85n.16) in our model of another person’s psychology – perhaps when we read the minds of others we begin by default attributing other psychological states as well, importantly many of our desires (as suggested by Goldman, 2006; Stereleny, forthcoming) and many of our valuings. Exactly the extent to

which default attribution occurs is a matter for further experimental exploration. The point is that applying this bit of theory to K&R-patterns might never have occurred to us had we not noted the similarities between these patterns and the data on egocentric biases.

Second seeing the existing K&R-patterns as instances of a more general pattern suggests further avenues of research. For example, there are results suggesting that people are more likely to display an egocentric bias in third-person mindreading when the target is judged to be attractive versus unattractive (e.g., Marks & Miller, 1982). Similar manipulations could be made in our studies. According to the current proposal, it would not be surprising to find that when a subject is provided with an attractive photo of, say, racist George, she will lean more heavily on her own psychology, thus exhibiting an even greater tendency to attribute the valuing of racial equality to him, than when presented with unattractive photo of George. To some extent, then, we should be able to manipulate the size of K&R-patterns. Exactly what those limits are is an interesting empirical question. The general point, again, is that by seeing the K&R-patterns so that they fit this more general body of results, further studies suggest themselves.

So it seems to me that the current evidence for claiming that MORAL GOODNESS is a feature of VALUING is not particularly overwhelming. The problem is that the empirical evidence reported here commits Knobe and Roedder to an implausible line on DESIRING. Of course what is generating their results and ours has not yet been settled. There are many options available, with plenty of opportunity for further investigation.

References

- Alexander, J. & Weinberg, J.M. (2007). Analytic epistemology and experimental philosophy. *Philosophy Compass*, 2, 56-80
- Brandt, R. (1979). *Theory of the good and the right*. Oxford: Clarendon Press.
- Bratman, M.E. (2000). Valuing and the will. *Philosophical Perspectives: Action and Freedom*, 14, 249-265.
- Fodor, J.A. (1998). *Concepts: Where cognitive science went wrong*. Oxford: Clarendon Press.
- Fumerton, R.A. (1983). The paradox of analysis. *Philosophy and Phenomenological Research*, 43, 477-497.
- Fumerton, R.A. (1995). *Metaepistemology and skepticism*. Lanham, MD: Rowman & Littlefield.
- Goldman, A.I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford: Oxford University Press.
- Goldman, A.I. (2007). Philosophical intuition: Their target, their source, and their epistemic status. *Grazer Philosophische Studien*, 74, 1-26.
- Goldman, A. & Pust, J. (1998). Philosophical theory and intuitional evidence. In M. DePaul & W. Ramsey (Eds), *Rethinking intuition: The psychology of intuition and its role in philosophical inquiry* (pp. 179-197). Lanham, MD: Rowman & Littlefield.
- Harman, G. (1993). Desired desires. In R.G. Frey & C.W. Morris (Eds), *Value, Welfare, and Morality* (pp. 137-157). Cambridge: Cambridge University Press.
- Holmes, D.S. (1968). Dimensions of projection. *Psychological Bulletin*, 85, 677-688.
- Kauppinen, A. (2006). *Lovers of the good: Comments on Knobe and Roedder*. Paper presented at the First Annual On-Line Philosophy Conference. Retrieved May 14, 2006, from <http://garnet.acns.fsu.edu/~tan02/OPC%20Week%20Three/Commentary%20on%20Knobe.pdf>
- Kauppinen, A. (2007). The fall and rise of experimental philosophy. *Philosophical Explorations*, 10, 95-118.
- Knobe, J. (2003). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology*, 16, 309-324.
- Knobe, J. (2007a). Experimental philosophy. *Philosophy Compass*, 2, 81-92.
- Knobe, J. (2007b). Experimental philosophy and philosophical significance. *Philosophical Explorations*, 10, 119-121.
- Knobe, J. & Roedder, E. (2006). *The concept of valuing: Experimental studies*. Paper presented at the First Annual On-Line Philosophy Conference. Retrieved May 14, 2006, from <http://garnet.acns.fsu.edu/~tan02/OPC%20Week%20Three/knobe.pdf>.
- Laurence, S. & Margolis, E. (1999). Concepts and cognitive science. In E. Margolis & S. Laurence (Eds), *Concepts: Core readings* (pp. 1-81). Cambridge, MA: MIT Press.
- Lewis, D. (1989). Dispositional theories of value. *Proceedings of the Aristotelian Society, Supplementary Volume*, 63, 113-137.
- Ludwig, K. (2007). The epistemology of thought experiments: First person versus third person approaches. *Midwest Studies in Philosophy*, 31, 128-159.
- Margolis, E. & Laurence, S. (2003). Concepts. In S.P. Stich & T.A. Warfield (Eds), *The Blackwell guide to philosophy of mind* (pp. 190-213). Malden, MA: Blackwell Publishing.

- Marks, G. & Miller, N. (1982). Target attractiveness as a mediator of assumed attitude similarity. *Personality and Social Psychology Bulletin*, 8, 728-735.
- Nahmias, E., Morris, S., Nadelhoffer, T. & Turner, J. (2005). Surveying freedom: Folk intuitions about free will and moral responsibility. *Philosophical Psychology*, 18, 561-584.
- Nichols, S. & Knobe, J. (forthcoming). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous*.
- Nichols, S. & Stich S.P. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Oxford: Clarendon Press.
- Osherson, D. & Smith, E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9, 35-58.
- Prinz, J. & Clark, A. (2004). Putting concepts to work: Some thoughts for the twenty-first century. *Mind & Language*, 19, 57-69.
- Ramsey, W. (1998). Prototypes and conceptual analysis. In M. DePaul & W. Ramsey (Eds), *Rethinking intuition: The psychology of intuition and its role in philosophical inquiry* (pp. 161-178). Lanham, MD: Rowman & Littlefield.
- Roedder, E. & Knobe, J. (2006). *Why not?* Paper presented at the First Annual On-Line Philosophy Conference. Retrieved May 14, 2006, from <http://garnet.acns.fsu.edu/~tan02/OPC%20Week%20Three/Reply%20by%20Knobe%20and%20Roedder.doc>.
- Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4, 328-350.
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B.B. Lloyd (Eds), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.
- Rosch, E. & Mervis, C.B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.
- Ross, L., Greene, D. & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13, 279-301.
- Savion, L. (2008). *Quick and dirty mental operations: The price of adaptive cognition*. Boston: Pearson Custom Publishing.
- Stanovich, K.E., Toplak, M.E. & West, R.F. (in press). The development of rational thought: A taxonomy of heuristics and biases. In R. Kail (Ed.), *Advances in child development and behavior* (Vol. 36). San Diego, CA: Academic Press.
- Stanovich, K.E. & West, R.F. (in press). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Psychology*.
- Sterelny, K. (forthcoming). The triumph of a reasonable man: Stich, mindreading, and nativism. In M. Bishop & D. Murphy (Eds.), *Stich and his critics*. Oxford: Blackwell Publishing.
- Swain, S., Alexander, J. & Weinberg, J.M. (forthcoming). The instability of philosophical intuitions: Running hot and cold on Truetemp. *Philosophy and Phenomenological Research*.
- Weinberg, J.M., Nichols, S. & Stich, S. (2001). Normativity and epistemic intuitions. *Philosophical Topics*, 29, 429-460.
- Whitbeck, L.B. & Gecas, V. (1988). Value attribution and value transmission between parents and children. *Journal of Marriage and Family*, 50, 829-840.

