

5 Experimental Approaches to Folk-Psychology: Moral Judgments and Pluralistic Accounts



danny.hammontree

(2006) *Thou Shalt Love Thy Neighbor* <http://www.flickr.com/photos/digitalgrace/90234275/>

Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported License.



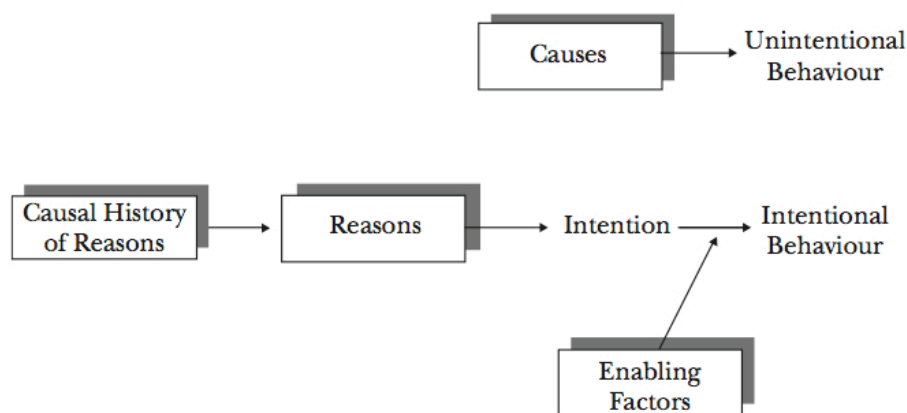
2008 Benoit Hardy-Vallée

www.hardyvallee.net/POSC

5.1 Folk-Psychology, From Intuitions to Experimentation

According to what we called the *standard account* in the last chapter, Folk-Psychology is a set of mechanisms or a body of knowledge that is involved principally in two cognitive operations: prediction and explanation. *Prediction* is an inference from mental states (beliefs-desires) to action, while *explanation* is an inference from action to mental states. This supposition—common in all major philosophy textbooks—has always been undisputed. Recently, however, experimental psychologist and philosophers undertake a more systematic approach of FP.

Malle and other social psychologists that try to decipher the folk concept of intentionality suppose three ‘layers’: a conceptual framework, a set of psychological processes and linguistic forms. The suggestion is that we all have a conceptual framework, akin to a “deep grammar” for social explanations. This framework is then expressed in linguistic forms by a set of psychological processes that govern the construction of explanations. Malle and his collaborators found that the conceptual framework, its psychological processing and its linguistic expression is relatively similar from one individual to another (Malle, 2001, 2007). First, almost everybody who agrees whether an action is intentional or not, rely preferentially on causes to explain the former and reasons to explain the latter. About 70% of the intentional actions are explained by primary reasons: beliefs, desires but also valuings (e.g. “she got home late because she liked the show”). When primary reasons are not evoked, subjects use either a causal history of reasons (CHR) explanation or an enabling-factor explanation. The first one explains why a person decided to do X not because of her beliefs/desires, but because of factors that bring about reasons to act: for instance, “she comes from a respectful culture”. The enabling-factor explanations cite—after the action is performed—the condition that made its performance possible without referring to the agent’s intentions or motivations (e.g. “she had two week to prepare the talk”). Here is how Malle et al. represent the folk concept of intentionality, based on their experiments:



(from Malle, 2007)

On the one hand, empirical research indicates that philosophers were right to see belief/desire as having a central role in FP (they are involved in the majority of explanations). On

the other hand, they may have overestimated the role of belief/desire attribution: folk-psychology seems to be a richer framework. Thus certain philosophers now propose a pluralistic account of FP.

5.2 Pluralism in Folk-Psychology

As Goldie (2007) argues, belief/desires explanations are rather thin: spelling out the mental content that explain an action is not really informative or explanatory. It is rather a re-description of the action. To rationalize an action (in the thin sense) consists just in showing how an action makes sense for someone, but there are many ways in which the action can make sense. One only has to look how we can make different, but all meaningful, predictions of someone's behavior.

Common-sense psychology, according to pluralist accounts (e.g. Andrews 2007, Goldie 2007) produce *thick*, rather than *thin*, explanations. Most of them are richer than beliefs/desires. A rapid search on Google using the expression "He did that because" output different styles of explanation:

He did that because of his friendship with the minister [valuing, causal history of reasons]

He did that because he loves you. [valuing, motive]

He did that because the consensus had moved [external factors]

He did that because he believes it and he did that because he wants to say it for her too. [beliefs-desires explanation]

He did that because of misconduct allegations [external factors]

He did that because I'm disabled [internal factors]

He did that because he had that quality of being genuine, of being authentic, of being unconditional and having a great zest for life [character trait]

He did that because he wanted the marriage to work [desire]

As these examples (and Malle's research) indicate, folk psychology is pluralistic and not limited to the attribution of propositional attitudes. We use many different concepts in explanation and prediction. We use factors such as **states** (depression, drunkenness, tiredness), **emotions** (she was angry, happy) **moods** (these days, he is tense, irritable, sad). Although these factors are not reasons (they are not located in the "space of reasons"), they do affect reasoning and thus can explain how someone decided to do something. We also use **stereotypes** (gender-, sex-, job-related), **rules** about what one should do given a certain situations or **generalizations** over past behaviors.

Motives are also often used in explanations. Motives explain an action by showing how something is desirable for the agent, how the object of desire is desirable. The motive does not need to be psychologically occurrent. One can be involved in a social practice without thinking about the motive (e.g. satisfying clients). Motive explanations are more general than B/D: they do not refer to "what's in someone's mind", but how a certain disposition is operative in a certain situation.

Another important category are personality **traits**. While motives indicate the effectiveness of disposition, traits are more general *dispositions*, or more precisely stable

properties that describe behavioral disposition. They are spontaneously used in descriptions, prediction, explanation, evaluation, interpretation, etc. Traits are thought to be less predictive than attitude (traits are biased: the fundamental attribution error biases people to neglect the influence of the situation on behavior) but we are generally biased in overestimating our predictive abilities and in confirming our hypothesis, so we don't notice faulty predictions (Andrews, 2007). Moreover, trait-based predictions are often local and contextualized, therefore we don't have many counterexamples (only when we generalize: we might be good at predicting a co-worker behavior at work, but not on a trip). Also, since the predictor is often part of the predicted situation, s/he can influence the predictability of the situation (e.g. the co-worker might always behave similarly with you). Traits are one heuristics among many others.

But is a trait just a complex of beliefs and desires? One might say that generous people have a desire to help others and belief that doing so-and-so will contribute to helping them. If that is the case, then trait-based predictions are just BD-predictions. But children use trait attributions only after they use BD attributions; hence trait attribution is a different capacity. Yet this does not show that this does not come from a generalization over BD attributions. However, although autistic children are unable to attribute mental states, they can learn (through Social Stories Therapy) to make predictions based on traits. The child can associate smiles with 'happy', and 'happy' with certain types of behavior. She can make prediction without attributing BD.

This 'folk theory of behavior' understanding is not restricted to autistic children; normal subjects also use it. Thus trait attributions are not reducible to belief attributions, and trait explanations are not mentalistic explanations in essence.

5.3 The Experimental Philosophy of Folk-Psychological Judgment

One of the consequences of the standard account, even if we consider the richer version provided by experimentalists, is that moral judgment allegedly takes place after folk-psychological judgment. First, we think “X does A”, then we judge that “X is blameworthy for doing A”. Experiments conducted by Joshua Knobe and other “experimental philosophers” (philosophers that use empirical method to evaluate philosophical or commonsense claims¹) showed that in many cases, people use a moral judgments to generate a FP judgment.

For instance, one of the first finding is that the moral status of a behavior can have an impact on whether or not people regard it as intentional (Knobe, 2007, p. 158). Take the following story

“The vice-president of a company went to the chairman of the board and said, ‘We are thinking of starting a new program. It will harm us increase profits, and it will also harm the environment.’ The chairman of the board answered, ‘I don’t care at all about harming the environment. I just want to make as much profit as I can. Let’s start the new program.’ They started the new program. Sure enough, the environment was harmed”

¹ The “X-Phi” community has a lively blog: <http://experimentalphilosophy.typepad.com/>

Most people claim that the chairman intentionally harmed the environment. If “harm” is replaced by “help”, however, people claim that the chairman unintentionally help the environment

Since the two cases have the same belief-desire structure, both actions should be seen as intentional, whether it is right or wrong. It turns out that in the “harm” version, most people (82%) say that the chairman intentionally harm the environment; in the “help” version, only 23% say that the chairman intentionally help the environment. This effect is called the “Knobe effect”, because it was discovered by philosopher Joshua Knobe. The experiment was also replicated with Hindi speakers—and led to similar results—and children. In the latter, the story was simplified. Children (2-,3-4-years old) were presented with a situation about a child A bringing a dog home, and knowing that it would make his friend B happy/upset. They were then asked “Does [A] make [B] happy/upset on purpose? When it upsets B, 4-years-old judge that A did it on purpose (but not younger children). This is a significant result, because this is the age where children usually acquire a full-blown theory of mind.

Other research led to similar results with the concept of valuing, that seems to be sensitive to moral consideration. Typically, philosophers and common sense distinguish desires from values. There is a difference between *wanting* a cigarette and *valuing* it: the happy smoker desires and wants it; the smoker who tries to quit smoking desires a cigarette but may not value it. It is hard, however, to sharply separate one from the other. Knobe suggested that valuing might be a prototype concept, that is, a concept that is not characterized by individually necessary and conjointly sufficient conditions (e.g., to be a grand-mother is to be the mother of the mother), but a prototype: *x* is *C* iff *x* has some of the features of a prototypical *C* (the typical car has 4 wheels, 4 places, etc.). Thus the prototype of valuing would be something like (i.e. a typical case of valuing will conform to most of these):

- 1) the agent has a conscious belief that *o* is good
- 2) the agent is motivated to promote *o*
- 3) the agent experiences guilt when she fails to promote *o* in circumstances where she could have
- 4) the agent has a second-order desire for *o* (i.e., a desire to desire *o*)

If there is a “Knobe effect” for valuing, we may be attributing valuations to people not only when we think they esteem something, but when we think that it is a morally good object. Maybe we consider that the heroin addict does not value heroin because it is really bad for him. If that is the case, then there is another feature of the valuing prototype:

- 5) whether the object *o* truly is morally good

In order to test this hypothesis, subjects were presented with a story about someone who has morally wrong desires, but seems to value something morally good. [Only condition 2 and 3 (motivation and guilt) are explicit].

“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.” [Knobe, 2007, p.161]

Then subject had to decide whether they agree to:

Despite his conscious beliefs, George actually values racial equality.

In another version, “racism” was replaced by “racial equality”. Thus subjects were presented with a story about someone who has morally good desires, but seems to value something morally good. Then subjects had to decide whether they agree to:

Despite his conscious beliefs, George actually values racial discrimination.

Results showed that when the attitude is good, it is considered to be one of the agent's values; when the attitude is wrong, it is not considered as one of the agent's values. This result provides some tentative support for the view that “moral judgments actually do play a role in people’s concept of valuing” (162)

Many other explanations of the Knobe effect were proposed: conversational pragmatics, emotional biases or factors that have nothing to do with the nature of the concepts. But for Knobe, they are not “the result of any kind of ‘bias’ or ‘distortion.’ Rather, moral considerations truly do figure in a fundamental way in the issues people are trying to resolve when they grapple with folk-psychological question” (Knobe 2007, p. 57). A pervasive feature of folk-psychology is that people’s intuitions about intentional actions “do appear to be influenced in some way by moral considerations, but it does not seem possible to pick out any particular sort of moral judgment that we are aware of making and say: ‘It is this sort of moral judgment that is influencing people’s intuitions about intentional action’.” (2007b, 101). The best explanation so far, according to Knobe and collaborators, would be that people make non-conscious moral judgments, and use them to attribute intentionality: “we simply determine whether the behavior in question violates any of the norms in the very limited set we are considering. If it does, we classify it as a transgression” (2007b, p. 102)

In a study by Pizzaro et al, subject were presented with stories about (1) gay men French-kissing on the street and (2) “couples” French-kissing; subjects in condition (1) considered that the couple were intentionally promoting gay kissing (supposedly because people have non-conscious preferences about gay sex).

All this could suggest that calling our everyday interpretative practice a “theory” of a folk version of scientific practice may be misleading for many reasons (the same reasons that could explain why many were reluctant think of FP as a moral-normative tool instead of a theory). There is a common implicit assumption in philosophy and cognitive science, according to which folk psychology is folk science, and science is not about moral claim, but prediction/explanation. Thus FP is not and could not be primarily about morality. For example, if a scientist changes her

prediction according to moral preferences, her judgment is considered distorted, biased, not profoundly moral. But is FP a commonsense, intuitive science?

Science as we know it, however, is a recent cultural phenomenon. There is little reason to suppose that these practices “reflect anything fundamental about our innate cognitive endowments” (2007, p. 164). Hence if FP is compared to a theory, it does not mean that it functions like a *scientific* theory. As Daniel Hutto argues, FP is holistic and structured like a theory, but FP is more a “framework” than a theory. A system of beliefs can be theory-like: it posits unobservable entities to explain observable ones, it goes “beyond the surface”. But this is something that religions do, yet religions and scientific theory are profoundly different. Even if FP is a theory, it has not the same “sensitivity to empirical evidence” (*Ibid.*, 166). Scientific method is, roughly about generating and testing hypotheses, and empirical evidence can be used to confirm or disprove the hypothesis. Theory-theorists argue that FP develops by falsifying mentalistic theories, nativists argue that there is no falsification, while others argue that it is an innate module parameterized during development (cf chapters 3 and 8). But sensitivity to empirical evidence is not specific only to scientific theories. Religious beliefs can be sensitive to empirical evidence, but no one claims that they are identical with scientific theories. The distinction lies mostly in what scientist do not or should not take into account. Scientists should consciously neglect moral considerations in adopting or revising a theory; but clearly these considerations are important in adopting or revising religious beliefs systems.

Another problem is that FP categories might not necessarily be best suited for scientific use (prediction-explanation); FP might create categories that serve other purposes. It may turn out “that the categories that prove most helpful in making moral judgments are completely orthogonal to the categories that prove most helpful in generating predictions and explanations.” (*Ibid.*168). Otherwise, we would need two systems: one for FP judgments (predictions), one for moral judgments (evaluations), plus rules for matching the two types of concepts. This is cognitively demanding, and unlikely. The “one-system that does both jobs adequately” solution does not work either: that would just show that FP is not like folk science. Science tries to develop precise, specific concepts tailored for prediction or evaluation, not general concept that does both job adequately. Hence folk theories are tools rather than predictive/explanatory framework (“theories”).

5.4 Folk-psychological accounts : a synthesis

We can now compare all the different conceptions of FP. Each account describes the basic entities that allegedly figure in the FP framework, the mechanism by which these entities are manipulated and the function of these operations.

	Entities	Psychological Mechanisms (operations)	Function
Standard Account	Beliefs & desires	(variable)	Prediction-explanations
TT	Beliefs & desires	Theoretical inference	Prediction-explanations
ST	Beliefs & desires	Simulation	Prediction-explanations
Gallager-Embodiment-Phenomenology	Sensorimotor capacities	Sensorimotor coordination	Interaction
Hutto-Narrative	Beliefs & desires	Narrative understanding and understanding narratives	Story-telling
Malle	Beliefs, desires, valuings, intentions, enabling factors, causal histories of reasons	Inferences	Prediction-explanation
Knobe	Beliefs & desires	Theoretical inference (but non-science-like)	Moral judgment, evaluation
Andrews-pluralism	Beliefs, desires, traits, emotions, moods, stereotypes, social norms, etc.	Inference	Prediction-explanations

5.5 References

- Andrews, K. *It's in Your Nature: A Pluralistic Folk Psychology*. Synthese.
- Goldie, P. (2007). *There Are Reasons and Reasons*. In D. Hutto & M. Ratcliffe (Eds.), *Folk Psychology Re-Assessed* (pp. 103-114). Dordrecht: Springer.
- Knobe, J. (2007). *Folk Psychology: Science and Morals*. In D. Hutto & M. Ratcliffe (Eds.), *Folk Psychology Re-Assessed* (pp. 115-130). Dordrecht: Springer.
- Knobe, J. (2007b). *Reason Explanation in Folk Psychology*. *Midwest Studies in Philosophy*, 31(1), 90-106.
- Nichols, S. (2004). *Folk Concepts and Intuitions: From Philosophy to Cognitive Science*. *Trends in Cognitive Sciences*, 8(11), 514-518.
- Pizarro, D., Knobe, J., and Bloom, P. 2007. "College Students Implicitly Judge Interracial Sex and Gay Sex to Be Morally Wrong" (unpublished manuscript, Cornell University)
- Folk Psychology Re-Assessed* (pp. 157-173). Dordrecht: Springer.

Thagard, P., & Kunda, Z. (1997). Making sense of people: Coherence mechanisms. In S. Read, & L. C. Miller (Eds.), *Connectionist models of social reasoning and social behavior*. Hillsdale, NJ: Erlbaum.