Philosophical Issues, 15, Normativity, 2005

MORAL PARTICULARISM AND TRANSDUCTION

Gilbert Harman
Princeton University

Introduction

There has been considerable recent discussion of one or another version of "moral particularism" as opposed to "moral generalism" (e.g., Dancy 1993, Sinnott-Armstrong 1999, Hooker and Little 2000, Kihlbom 2002, Väyrynen 2004). Moral generalism holds, either that specific moral truths have their source in general moral principles, or that reasonable or justified moral decisions and beliefs are based on the acceptance of general moral principles. Moral particularism rejects moral generalism.

There are stronger and weaker versions of moral generalism and correspondingly weaker and stronger versions of moral particularism. *Strong moral generalism* holds, either that *all* specific moral truths have their source in general moral principles, or that *all* reasonable or justified moral decisions and beliefs are based on the acceptance of general principles. *Weak moral generalism* holds, either that at least *some* moral truths have their source in general moral principles, or that at least *some* reasonable or justified moral decisions and beliefs are based on the acceptance of general moral principles. Since the denial of a strong claim is weak and the denial of a weak claim is strong, strong moral particularism denies weak moral generalism and weak moral particularism denies strong moral generalism.

I will be concerned with one of the issues between strong moral generalism and weak moral particularism.

The issues between particularism and generalism are metaphysical, to the extent that they concern the source of moral truths, and epistemic (or psychological), to the extent that they concern the source of reasonable or justified moral decisions and beliefs.

The metaphysical and epistemic issues are possibly connected. Metaphysical moral particularism is often defended by appeal to such theses as that for an act to be morally wrong is for it to be such as to be seen or judged wrong by a competent moral judge (Wiggins 1998, McDowell 1998), given that epistemic moral particularism applies to the competent moral judge.

In what follows, I am going to concentrate on the epistemic issue. That is, I am going to be concerned with the question whether all reasonable or justified moral decisions or beliefs have to be based on general moral principles. Furthermore, I am going to be concerned with an issue about the principles and processes involved in arriving at reasonable or justified moral decisions or beliefs and not for example with principles offered later in justification or defense of a moral decision or belief.

Two important and difficult issues immediately arise. First, what *counts* as a general principle? And, second, what is it for a moral decision or belief to be *based on* a general moral principle?

Moral Principles

One issue about the relevant moral principles is whether they are to be conceived as exceptionless rules or as default principles. A related issue is whether the relevant principles are expressible in language and explicitly appealed to when people defend their actions or beliefs.

The issues are related because the principles to which people explicitly appeal tend to be principles they think have exceptions—principles like "It is wrong to break a promise" and "It is wrong to kill another person." The average person is not normally able to formulate *any* exceptionless moral principles, although some philosophers claim to be able to do so. Consequently, many participants in the particularism-generalism debate have supposed that the relevant principles are prima facie or default principles rather than exceptionless rules.

The trouble with this argument is that not only are people not normally able to formulate exceptionless moral principles, they often have trouble formulating even maxims or default principles that would account for some of their seemingly reasonable moral judgments. For example, ordinary people judge that it is morally right or permissible to turn a trolley onto a side track in order to reduce the number of people killed from five to one, but it is not in the same way right or permissible to cut up a hospital visitor in order to distribute organs to patients who need them to live, even if this will reduce the number of people who die from five to one (Foot 1978) and these judgments are not obviously unreasonable. Most people judge that it is wrong to push a fat man off a bridge so that a runaway trolley will hit him and stop before running over the five people trapped ahead, but permissible

to divert a missile heading for New York City so that it explodes instead over Worcester, Massachusetts, where fewer people will be killed (Thomson 1986). Most people think it is worse to refuse to give food to a beggar on the grounds that bodies are needed for medical research than to refuse to give food to the beggar simply out of inertia because one cannot be bothered (Foot 1978). People think it is right for someone to give benefits to his or her children in preference to other children but wrong for a brother and sister to engage in sexual intercourse even if they use contraception (Haidt, 2001). Although people make such judgments and often seem reasonable to do so, they often appear to be quite incapable of stating principles, even prima facie or default principles, that might explain them.

This by itself does not immediately establish moral particularism. Hare (1952, pp. 56–78) points out that in many instances people are sensitive to relevant aspects of a situation and respond in ways that may indicate at least an *implicit* acceptance of general exceptionless principles. According to Hare people often follow certain exceptionless principles that capture what they have learned *to do* but may not be able to express in words. Their understanding of the principles may be practical rather than theoretical, know-how rather than knowledge that. So, according to Hare, people do not have to be able to formulate in words the principles on which they act.

I want to count Hare's view as an example of moral generalism, so I want to allow that people can act or judge on the basis of complicated exceptionless general principles when their understanding of the principles is purely practical in this sense, that is, when they act from a disposition of the relevant sort. Someone might in this way act or judge on the basis of a complicated exceptionless principle that cannot even be formulated at all in language.

The relevant sort of principle is not to be identified with any sort of linguistic expression. Instead, let us identify such principles with partial functions in the mathematical sense of function—partial functions that map epistemic descriptions of cases into decisions or judgments about the cases.

This is an appropriate way to think of principles with respect to the epistemic issue. With respect to the metaphysical issue, the relevant moral principles might be identified with functions from nonmorally specified possible situations or cases in nonmorally specified possible worlds to moral assessments of the situations or cases in those worlds. So understood moral generalism would seem to be equivalent to the claim that moral truths supervene on nonmoral truths.

To return to the epistemic issue. What is it for a decision or judgment to be "based on" a principle *P*, conceived as a function in this sense? Following Hare, we might suppose that, *to a first approximation*, a decision or judgment is *based on* a principle *P* if it arises out of a *settled disposition* to make decisions or judgments specified by the function *P*.

But your current implicit moral principles cannot simply be *identified* with your current settled dispositions to reach moral verdicts given epistemic situations, because you might for example have a settled disposition to reconsider your principles given the actual occurrence of certain epistemic situations.

For example, your principles might imply that, given epistemic situation *E* it would be wrong to *D*, where you think *E* is highly unlikely to occur. Your settled dispositions might also be such that, if *E* were to occur, you would reconsider your principles and change your view about this sort of case.

More generally, given the actual occurrence of certain unexpected epistemic situations, you might be disposed to come up with particular verdicts not covered by your present principles. Furthermore, your disposition to come up with those particular verdicts might be based on a general principle of inference or reasoning, without your having already accepted a moral principle or any principle in the relevant sense about those cases. (I will come back to this point in discussing "transduction" below.)

So, the connection between your implicit principles and your dispositions to make decisions or judgments is somewhat unclear.

Perceptual Analogy?

Some defenders of particularism appeal to a perceptual analogy (Wiggins 1998, McDowell 1998). Given your background beliefs and awareness of the facts of the case, you immediately "see" that something is wrong without basing this conclusion on a general principle that implies it. To be sure, in accepting that conclusion you are committed by universalizability to the general principle, "Anything exactly like this is wrong." But your acceptance of the moral conclusion about the case is not based on *prior* acceptance of this general principle.

This might illustrate moral particularism, but also it might not. The fact that such perception-like processes seem to be immediate does not prevent them from involving the application of an exceptionless rule in the relevant sense. But if the perceptual analogy does not support weak epistemic moral particularism, what could support it?

Testimony

Pekka Väyrynen (personal communication) notes that someone might justifiably act or form a justified moral belief about a particular case on the basis of the testimony of a trusted adviser. It might be that the agent acts on the basis of a general moral principle that one should appeal to this sort of advisor for situations of this sort (Sartre 1956, p. 297), but the agent might

simply be relying on the nonmoral assumption that this particular adviser is generally reliable, as Elizabeth Harman forcefully pointed out to me. Justified reliance on testimony in this way should not be enough to establish weak epistemic moral particularism, so let us ask whether there are other ways in which a justified moral decision or moral belief fails to rest on prior acceptance of a moral principle.

Moral Particularism and Transduction

Suppose you are faced with an epistemic situation *E* that is not covered by principles you already accept, you cannot rely on the opinion of a trusted adviser, and you want to reach a verdict about the case. Strong epistemic moral generalism says the rational way to proceed is first to reason to a general principle and then use that principle to settle the present case. Weak epistemic moral particularism says it can be rational directly to reach a verdict on the present case without basing that verdict on a prior acceptance of a general principle. Particularism can allow that your verdict may (given other things you accept) *commit* you to some general principle (e.g., via a principle of universalizability), even if your verdict is not *based on* your acceptance of that principle.

This issue between moral generalism and moral particularism is analogous to (or an instance of) the question whether there can be a *direct inductive inference*—also called a *transductive inference* (Vapnik 1998; 2000, p. 293). A direct or transductive inference goes from data about previous cases to a classification of a new case. This classification does not involve first making an inductive generalization from data about previous cases and background assumptions, maybe using *inference to the best explanation*, and then deducing a conclusion about the new case from that inductive generalization.

Vladimir Vapnik and others have developed methods of transduction that in certain cases give better results than inductive methods that infer general principles (Joachims 1999, Vapnik 2000, Weston et al. 2003, Goutte et al. 2004). And, if transduction is sometimes more reasonable in nonmoral reasoning, perhaps it can also be more reasonable in moral reasoning. In that case weak epistemic moral particularism would be supported over strong epistemic moral generalism.

The issue is complicated by the fact that ordinary methods of inductive generalization are not limited to the methods Vapnik counts as "inductive methods" in statistical learning theory. Nevertheless, a careful study supports the claim that transduction can do better than ordinary inductive generalization in certain cases. In order to explain this, I need to say something about some of the methods studied in statistical learning theory (Hastie et al. 2001).

Statistical Learning Theory

Suppose you want to be able to classify various instances on the basis of N observed features, features like color, size, temperature, etc. For example, you might want to classify leaves, or handwritten symbols, or viruses. Each feature of an instance can take a variety of values, which we might suppose can be represented numerically. Suppose there is an N-dimensional "feature space," with separate dimensions for each of the N features, the color dimension, the size dimension, the temperature dimension, etc., where a point in the feature space represents a possible cluster of observed features.

A *classifier* is a system that takes a representation of a cluster of observed features and outputs a classification. (In the simplest case the classification has two possible values, YES and NO.) Methods of learning are methods for using data in order to train a classifier to do well on new cases.

A *perceptron* is a simple classifier that has N weighted inputs, one for each of the N observable features an object can have. The classifier takes a weighted sum of the feature values and outputs YES if this sum is greater than 0 and NO if it is not.

The perceptron is trained using data. Before learning, the weights on the inputs are assigned random values (positive and negative). Then, for each datum, the values of its features are input to the perceptron. If the perceptron outputs the correct classification of that datum, no change is made in the perceptron. If the perceptron outputs the wrong classification, the weights on the inputs are changed slightly in ways that would make the weighted sum of the inputs closer to the desired value. This procedure is repeated going through the data many times.

It can be shown that this procedure will eventually yield a perceptron that does well at classifying all the data, as long as the data can all be correctly classified by some perceptron. Once the data are correctly classified, the perceptron can be used to classify new cases. There are theorems about how good to expect results to be on new cases.

Unfortunately, a perceptron can only represent classifications that are linearly separable. Famously, a perceptron cannot represent the XOR classification. Given two inputs, X and Y, an XOR classifier indicates YES if the product $XY \leq 0$ and NO if the product $XY > 0$. A perceptron cannot do this, because the YES region cannot be separated from the NO region by a straight line.

Feed-forward neural networks address this problem by combining several "layers" of perceptrons, the outputs from earlier layers serving as inputs to the perceptrons in later layers. A simple two layer feed-forward neural network can represent XOR and it turns out that any rule of classification can be approximated by some three layer network. Given also a small

change in the calculation performed by each node of the network, there is a learning rule of back-propagation of error that often works well, although to do well the method may need a great deal of data. Furthermore, it sometimes gets stuck in "local minima."

Support vector machines (SVMs) represent an alternative response to the limitations of perceptrons. Instead of adding additional layers to the network, SVMs begin by mapping the data into a higher-dimensional space in which the data can be linearly separated. So, for example, the data in the XOR problem might be mapped into a three dimensional space in such a way that each point (X,Y) is mapped onto (X,Y,Z), where $Z = XY$. The transformed data points are linearly separable in that 3-space by the plane that is perpendicular to the Z axis at $Z = 0$, i.e., the plane defined by the X and Y axes, because the YES items are below that place and the NO items are above it.

Usually, in practice, many different hyperplanes will separate the data in the transformed higher-dimensional space. In one SVM approach, the hyperplane that *maximally* separates the data (as represented by "support vectors") is chosen to classify all cases as on one side (YES) or the other (NO) of the chosen hyperplane.

Transduction

These learning methods all involve learning total classifications. Perceptrons, multi-layer feed-forward networks, and standard SVMs all yield classifications of every possible set of features. This is also true for other methods, such as nearest neighbor methods, where the data are placed at various points in a feature space for which a distance measure has been defined and where a new item is assigned the same classification as the nearest datum.

Vapnik's idea (1998, 2000) about transduction is to modify SVMs so as not to yield such total classifications. In particular, transduction does not use prior data and background assumptions to choose a single linear separation in the transformed higher-dimensional space. For one thing, transduction also uses the information that certain new cases have come up to be classified. Then instead of trying to select a single best separation, transduction selects a subset of separations that (a) correctly classify the data and (b) agree on their classifications of the new cases. In one version, the selected separations also (c) disagree as much as possible on the classifications of other possible cases.

A related version of transduction uses not only the information that certain new cases have come up to be classified but also the information that there is a certain set U of cases ("universum") that are taken to be hard to classify. In this version, transduction selects the subset of linear separations

satisfying (a) and (b) but disagreeing as much as possible on the classification of the hard cases in U.

Transduction performs considerably better than other methods in certain difficult real-life situations involving high-dimensional feature spaces where there is relatively little data (Joachims 1999, Weston et al. 2003, Goutte et al. 2004).

Transduction and Induction

Vapnik (2000, p. 293) claims that transduction is a kind of inference in which the classification of new cases is not based on a prior inductive generalization. What can we say about this claim?

One obvious point is that ordinary inductive generalization does not normally arrive at a total classification, so the mere fact that transduction fails to do so does not distinguish it from ordinary inductive generalization. A second point is that nothing prevents ordinary inductive generalization from making use of the information that certain new cases have come up to be assessed and that certain other cases are going to be hard to classify, so the fact that transduction makes use of such assumptions also does not distinguish it from ordinary inductive generalization.

A third point is that transduction involves the implicit acceptance of a non-total principle P, corresponding to the selected subset of separations in the transformed higher-dimensional space. To be sure, P is not previously accepted, so the classification of the new cases is not determined by a *previously accepted* principle. But that is true whenever we reason to new principles.

On the other hand, there is an important sense in which the acceptance of P does not yield any additional classifications beyond the classifications C gives to the new cases on this occasion. P is determined entirely by prior commitments plus the classifications given to the new cases. Before the new cases arose, a principle Q is accepted that is equivalent to a set of separations of the data in the transformed higher-dimensional space. Transduction leads to classifications C of the new cases. The new principle P follows from C and Q. More precisely, $Q \rightarrow (C \leftrightarrow P)$. Given Q, to accept the classifications C is equivalent to accepting the principle P. To accept the one is to accept the other. So, the acceptance of one cannot be based on the *prior* acceptance of the other. Vapnik is right; transduction leads to a conclusion that is not based on a prior inductive generalization.

So, *moral* transduction, if it occurs, leads to the acceptance of a specific moral conclusion that is not based on the prior acceptance of a moral principle that covers that case. Acceptance of the specific moral conclusion involves but is not based on acceptance of a new moral principle.

Of course, as noted earlier, although acceptance of the specific moral conclusion would in this case be based on a general principle of transduction, a general principle of reasoning like transduction is not the sort of general *moral* principle that is at issue between moral generalism and moral particularism.

It would seem that moral transduction could be a reasonable form of direct inference. We can now ask whether it actually ever occurs. How could we tell whether people ever use moral transduction or, indeed, any sort of transduction?

This is a difficult issue. We cannot simply appeal to the fact that people seem to reach reasonable moral conclusions without being able to state moral principles, because the relevant principles do not have to be conscious or even formulable in language. We cannot appeal to the fact that in some cases we seem to "see" that something is right or wrong, because such moral perception can involve the application of exceptionless principles that determined by our neural circuitry.

Perhaps this issue could be resolved through the discovery of the relevant sort of SVM type processing in the brain. But I do not have anything interesting to contribute to the discussion of such a possibility.

Objections to Moral Statistical Learning Theory

I want to conclude by briefly noting some possible objections to supposing that statistical learning theory could be relevant to the issues between epistemic moral generalism and epistemic moral particularism.

Objection: Statistical learning theory is concerned with reasoning to a classification of a new case given data consisting in information about the classification of other cases. But moral reasoning is not reasoning about how to classify something, it is practical reasoning about what to do.

Reply: Reasoning as to whether a certain act is wrong is reasoning as to whether to classify the act as wrong. Similarly, reasoning as to whether to do a certain act is reasoning as to whether to classify the act as something to do.

Objection: Moral reasoning is not reasoning from data.

Reply: Sometimes in moral reasoning people compare the present case with other cases which they (think they) know how to classify as right or wrong. That counts as reasoning from data.

Objection: Statistical learning theory is concerned with classifications that can be correct or incorrect, but moral judgments are merely expressions of feelings or commitments that cannot be correct or incorrect.

Reply: People reason as if moral judgments can be correct or incorrect, so it is appropriate for them to reason in accordance with principles that do well in getting correct conclusions from correct data.

Objection: Statistical learning theory takes an "instrumentalist" approach to a subject, seeking to minimize expected error in the classifications of new cases. The history of science shows that this is a shortsighted approach and that it is better to adopt a realist approach that attempts to find the underlying explanatory principles behind the phenomena.

Reply: The history of science and technology indicates that sometimes basic research pays off and sometimes it's better to undertake more applied research. Sometimes it pays to try to find the hidden principles underlying the phenomena, and sometimes that is too difficult and it pays to stay with more instrumentalist approaches.

Objection: But what we really want to know are the basic underlying principles, not the superficial appearances.

Reply: Sometimes we want one and sometimes we want the other.

Objection: Moral conclusions anyway should be based on an understanding of why they are true. So moral reasoning at least cannot accept the instrumentalist outlook behind statistical learning theory.

Reply: The objection is based on a "Platonic" conception of morality as concerned with something like a moral law just as physics is sometimes conceived as concerned with finding basic underlying laws. Even given that (controversial) conception of morality, it must be conceded that sometimes people can make reasonable moral judgments without having much of an understanding of the basic underlying principles.

Furthermore, there are other conceptions of morality, for example, as something like law but more informal, resting on understandings among people and the ways in which they assess various cases.

It is interesting that legal reasoning in hard cases can reflect aspects of the second version of transduction mentioned above. Appellate justices in the United States often tend to try to decide only the case before them without trying to decide various other possible hard cases, including possible cases that are in some respects similar to the case being decided. Courts often try to decide a very narrow issue. Occasionally members of the US Supreme Court disagree about whether the majority decision in a given case has implications for various other possible cases, with those in the majority arguing that it has no such implications. On such occasions it is taken to be

a virtue in a decision that the decision leaves other possible nearby cases maximally undecided.

For example, in their 2003 decision in Lawrence et al. *v* Texas, the U. S. Supreme Court majority ruled that Texas' anti-sodomy law unconstitutionally discriminated against homosexuals. In a concurring opinion, Justice Sandra Day O'Connor argues that the rationale of the court's decision would not also apply to laws prohibiting same sex marriages. In dissent, Justice Antonin Scalia disagrees. Both opinions take for granted that it is undesirable to decide the issue then under consideration in a way that would also decide the other issue.

I should say that the analogy with appellate legal decisions is potentially misleading. The legal decisions set precedents, but transductive inferences need not set precedents. That is, the conclusions of transduction need not be treated as new data for further reasoning. And those conclusions might be retracted as further cases arise that indicate more about the background probability distribution.

Conclusion

This paper has been concerned with the issue whether someone can be reasonable or justified in accepting a specific moral judgment not based on the prior acceptance of a general exceptionless moral principle, where acceptance of a general principle might be tacit or implicit in and might not be expressible in language. The issue is an instance of a wider issue about direct or transductive inference. Developments in statistical learning theory show that such an inference can be more effective than alternative methods using inductive generalization. Such transduction can therefore be reasonable. This result carries over to moral transduction, although it is a difficult empirical issue whether people actually engage in any sort of transduction, including moral transduction.

Acknowledgments

Walter Sinnott-Armstrong and Pekka Väyrynen gave me useful comments on earlier drafts of this paper. I am indebted to discussions with Vladimir Vapnik about transduction. My account of statistical learning theory is based on a textbook Sanjeev Kulkarni and I are writing about *Epistemology and Learning Theory*, in connection with an introductory course we give at Princeton University with that title. I have received useful help from audiences at Princeton University, at an Empirical Moral Philosophy workshop at Dartmouth, and at a meeting of SOFIA in Huatulco, Mexico, at which Barbara Herman and Martin Stone were commentators.

References

Dancy, J. (1993) *Moral Reasons*. Oxford, Blackwell.

Foot, Philippa. (1978) "The Problem of Abortion and the Doctrine of the Double Effect," in *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford, England: Blackwell.

Goutte, C., Cancedda, N., Gaussier, E., Déjean, H. (2004) "Generative vs Discriminative Approaches to Entity Extraction from Label Deficient Data." JADT 2004, 7es Journées internationales d'Analyse statistique des Données Textuelles, Louvain-la-Neuve, Belgium, 10–12 mars.

Hooker, B., and Little, M. (eds.). (2000) *Moral Particularism*. New York: Oxford University Press.

Kihlbom, U. (2002) *Ethical Particularism*. Stockhold Studies in Philosophy 23. Stockholm: Almqvist and Wiksell.

Korsgaard, C. (1996) *The Sources of Normativity*. Cambridge, Cambridge University Press.

Haidt, J. (2001) "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* **108**, pp. 814–34.

Hare, R. M. (1952) *Language of Morals*. Oxford: Oxford University Press.

Hastie, T., Tibshirani, R., and Friedman, J. (2001) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.

Joachims, T. (1999) "Transductive Inference for Text Classification Using Support Vector Machines." In I. Bratko and S. Dzeroski, editors, *Proceedings of the 16th International Conference on Machine Learning*, pp. 200–9. San Francisco: Morgan Kaufmann.

McDowell, J. (1998) *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press.

Scanlon, T. (1998) *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Sinnott-Armstrong, W. (1999) "Varieties of Particularism." *Metaphilosophy* **30**, pp. 1–12.

Sartre, J. P. (1956) "Existentialism Is a Humanism." In W. Kaufmann, editor, *Existentialism from Dostoevsky to Sartre*. New York: Meridian, pp. 287–311.

Thomson, Judith Jarvis. (1986) "Killing, Letting Die, and the Trolley Problem." In W. Parent, editor, *Rights, Restitution, and Risk: Essays in Moral Theory*. Cambridge, MA: Harvard University Press, pp. 78–93.

Vapnik, Vladimir. (1998) *Statistical Learning Theory*. New York: Wiley.

Vapnik, Vladimir. (2000) *The Nature of Statistical Learning Theory*, second edition. New York: Springer.

Väyrynen, P. (2004) "Particularism and Default Reasons." *Ethical Theory and Moral Practice* **7**, pp. 53–79.

Weston, J., Pérez-Cruz, F., Bousquet, O., Chapelle, O., Elisseeff, A., and Schölkopf, B. (2003) "KDD Cup 2001 Data Analysis: Prediction of Molecular Bioactivity for Drug Design—Binding to Thrombin." *Bioinformatics*.

Wiggins, D. (1998) *Needs, Values, and Truth*, 3rd edition. Oxford: Oxford University Press.