

## REASONS

GILBERT HARMAN  
Princeton University

### I

A number of ethical theories take the correctness of a moral "ought" judgment to depend on how that judgment is related to certain practical principles, although the theories differ concerning what they take the relevant principles to be. Kant, for example, says that the relevant principles are the universal principles of pure practical reason. Hare says that they are whatever universalizable principles that the person making the "ought" judgment subscribes to. Sartre says that they are the principles that are accepted by the person about whom the "ought" judgment is made. The convention theory says that the relevant principles are those conventionally accepted by the members of some contextually indicated group that includes both the person being judged and the person making the judgment.

Theories of this sort can also differ in what they say about the required relation between practical principles and moral judgment. For example, Hare's theory is that this is a logical relation that is to be explained within a logic of imperatives. An alternative and more plausible idea is that the relevant relation is a weaker nonlogical relation of the sort that holds between reasons and the thing for which the reasons are reasons.

According to Hare, the relevant principles can be formulated as general imperatives; and an "ought" judgment is correct, in relation to a set of general imperatives, if a corresponding imperative follows logically from the set of

general imperative taken together with true factual assumptions. For example, suppose that the person making the judgment accepts the principle, "Anyone, give someone what you owe him if you can and he asks for it back"! Suppose also that the facts include (1) that Jones owes Smith ten dollars and (2) that Smith has asked for the money back. Hare believes that from these suppositions we can derive the imperative "Jones, pay Smith the ten dollars you owe him"! So, in Hare's view, the judgment that Jones ought to pay Smith the ten dollars he owes him is correct, in relation to that general principle, given those facts. This idea, about the required relation between particular "ought" judgment and general principles, could also be accepted in a Kantian, Sartrean, or convention theory morality. Any of these theories could say that a moral "ought" judgment is correct if, and only if, it is a logical consequence of the relevant practical principles and the facts of the case. The difference between the theories would then have to do with the nature of the relevant principles, and not with the required relation between principles and particular "ought" judgments.

However, the idea that there must be this sort of logical relation between principles and particular judgments is not very plausible. For one thing, it implies that the relevant principles are so detailed that they determine logically exactly what we are to do in every possible circumstance. The idea that moral principles have this sort of precision has an air of unreality.

There is, furthermore, a looseness to "ought" judgments that this idea fails to capture. "*P* ought to do *D*" is not the only sort of judgment that we want to make in relation to the relevant practical principles. We also want to be able to say such things as, "It would be good of *P* to do *D*", "It would be wrong of *P* not to do *D*", "*P* might do *D*", "*P* must do *D*". These judgments do not all mean the same thing. Some are appropriate where others are not. Hare's account seems on its face more appropriate for a judgment like "*P*

must do *D*" than for "*P* ought to do *D*". For, to say that *P* ought to do *D* is not necessarily to say that *P*'s doing *D* is absolutely required.

It would be better, if that is what is meant, to say that *P* has to do *D* or that it would be wrong of *P* not to do *D*. We can suppose that *P* ought to do *D* without supposing that he absolutely must do *D* or even that it would be wrong of him not to do *D*. Saying that *P* ought to do *D*, on the other hand, is a stronger thing to say than saying simply that it would be good of *P* if he were to do *D*.

## II

This suggests an alternative account according to which the relevant practical principles are vague and lack the sort of legalistic precision that would be required for Hare's account to work. In this alternative theory, the relevant principles commit us not so much to particular actions in particular circumstances as to certain general aims and goals. We are to respect others; we are to try not to harm them; if there is no great cost to ourselves involved, we are to try to help those who need help; and so forth. There are also principles regarding duties and divisions of responsibilities, e.g. within families — parents are to be responsible for the education and well-being of their children and so forth. In this view, the relevant moral principles are too vague to determine precisely what someone ought to do, if this has to be determined logically in the way that Hare suggests, since the relevant principles do not have the sort of precision necessary for us to be able to deduce what ought to be done, given the facts.

Still in this alternative view, as in Hare's view, to accept principles as practical principles is to be motivated to act in certain ways. It is to have certain goals and ends in addition to the goals and ends you would have if you had not accepted those principles. And, just as your other goals, ends, desires, and plans can give you reasons to do things, so too can the

goals and ends that you have e.g. as the result of accepting the conventions of society. But, to say that these principles give you reasons to do things is not just to say that these principles logically imply certain imperatives. The reason relation is weaker than, or at least different from, the relation of logical implication via a logic of imperatives. It is this vaguer relation of reasons that, in this view, correctly expresses the connection between the relevant practical principles and particular "ought" judgments that are correct, in relation to those principles, given the facts.

In this view, then, to say that *P* ought to do *D* is to say that *P* has sufficient reasons to do *D* that are stronger than reasons he has to do something else. If what you mean is that *P* morally ought to do *D*, you mean that *P* has sufficient moral reasons to do *D* that are stronger than the reasons he has to do something else. In other words, given the relevant practical principles, for example the principles that *P* accepts as the conventions of society, *P* has sufficient reasons deriving from those principles for doing *D*, etc.

The relevant reasons can be of varying strengths, which accounts for the various sorts of judgment that might be made: "*P* ought to do *D*", "*D* is the best thing *P* can do", "*P* must do *D*", "*P* may do *D*", and so forth. For example, given that *P* intends to adhere to certain principles and given the facts, it may be that there are good reasons for *P* to do *D* yet it would not all be a mistake for *P* not to do *D*. In that case, it may not be true that *P* positively ought to do *D* although it would be good of *P* to do *D*. On the other hand, if given *P*'s intention to adhere to those principles and given the facts, it would be a mistake for *P* not to do *D*, *P* ought to do *D*. And, if it would be irrational for *P* not to do *D*, given that he really does intend to adhere to those principles, he must do *D*. Not doing *D* in such a case would be incompatible with continuing to intend to adhere to those principles. To say, in relation to certain principles, that *P* must do *D* is to say that, if *P* does not do *D*, that will show that he does not

in fact seriously intend to adhere to those principles. Finally, to say that *P* may do *D* or might do *D* in relation to certain practical principles is to say that *P*'s acceptance of those principles does not make it irrational for him to do *D*.

A theory of this sort is sometimes called a "good-reasons" analysis of moral "ought" judgments and related moral judgments. In this view, such judgments say something about the moral reasons a person has to do things. A similar analysis would be inappropriate for other sorts of moral judgment, such as the judgment that *P* is morally evil in doing *D*, since such judgments do not depend in the same way on assumptions about *P*'s reasons to do things.

Good-reasons analyses can also be given for other senses of "ought" and related words like "may" and "must". Recall that the word "ought" appears to have at least four different meanings. In addition to the moral "ought" that we have been discussing, there is the simple "ought" of rationality, as when we say that the bank robber ought to use the rear door; the evaluative "ought", as in "There ought to be more love in the world"; and the "ought" of expectation, as in "The train ought to be here in three minutes." Corresponding to these different senses of "ought" are different senses of "must" and "may". There is, for example, a "must" of expectation, as when we say, "The train must have arrived by now", and there is a "may" or "might" of expectation, as when we say, "The train may arrive in few minutes", or "The train might have arrived at noon; I'm not sure". Similarly, there are a "must", a "may", and a "might" of evaluation and also of simple rationality. We say, for example, "The hurricane mustn't hit Miami" or "The bank robber must cut the alarm wires if he is to escape detection."

Good-reasons analyses, suitably modified, are appropriate to these different cases. In the case of the "ought", "must" and "may" of expectation, what is relevant are reasons to believe things. To say that the train ought to be here soon is to say that there are good reasons to think the train will be

here soon. To say that the train must have arrived by noon is to say that the reasons for thinking this are conclusive. To say that the train may not have arrived at all is to say that it is not incompatible with the reasons we currently have to suppose that the train has not arrived at all. Similarly, to say that the hurricane mustn't hit Miami is to say that there are overwhelming reasons to hope that the hurricane will not hit Miami. To say that there ought to be more love in the world is to say that there are reasons for wishing that there were more love in the world. To say that the bank robber ought to use the back door is to say that he has good reasons to do that. To say that he must cut the alarm wires is to say that it would be irrational for him not to do so, given his ends and given the facts.

### III

One advantage of such a good-reasons analysis is that it allows us to account in this way for various uses of "ought", "may" and "must". Another is that the good-reasons analysis helps to explicate an aspect of our use of "ought" that was emphasized by W. D. Ross, who argued that there were two moral "oughts", a prima-facie "ought" and an all-things-considered "ought". Ross argued that, if we try to state our moral principles using the word "ought" we must use what he calls the prima-facie "ought". For example, we say, "You ought to keep your promises", "You ought to tell the truth", "You ought not to injure others", "You ought to help those in need", and so forth. But we do not suppose that these principles are absolute. We do not suppose that you ought to keep every promise; we allow that there are circumstances in which you may break a promise — indeed there are circumstances in which you ought to break a promise. The same is true of other moral principles. "All moral principles have exceptions".

Hare's reaction to this point is to suppose that our moral principles are really much more complicated. For Hare, the

"exceptions" are built into the moral principles. But Ross took a different tack. According to him, a principle like "You ought to keep your promises" is true as it stands; but this principle does not mean that there are no situations in which you may break your promise. According to Ross, the "ought" used in stating this principle is the prima-facie "ought". What the principle says is that you ought, prima-facie, to keep your promises. More precisely, it means that, if you have promised to do something, that gives you a moral reason to do what you have promised to do. If you have no other reasons, then you should do what you have promised to do. If you have other reasons, then you must weigh your various reasons against each other in order to decide what you ought to do all things considered.

For example, you have promised to attend a meeting but your aunt has just died so you also have an obligation to attend the funeral. You therefore have conflicting obligations. Given that you have promised to go to the meeting, you ought (prima-facie) to go to the meeting. Given that your aunt has died and your parents will be expecting you at the funeral, you ought (prima-facie) to go to the funeral. These "ought" statements do not contradict each other, since they are prima-facie "ought" statements. Neither statement says what you ought to do all-things-considered. This can be determined only by weighing your reasons against each other and deciding which is stronger.

Ross' theory of the prima-facie "ought" makes a great deal of sense from the point of view of the good reasons analysis. In this view, moral principles indicate the sorts of things that you have moral reasons to do. What you should do all-things-considered, is not determined by some further moral principle but is determined in whatever way conflicting reasons are weighed against each other, be they moral reasons or reasons of other sorts.

Putting this in another way, we might take Ross' theory to be a remark about the logical form of "ought" statements.

Strictly speaking, we can say, an "ought" statement has the form, "Given *C* (i. e. in relation to the fact that *C*), *P* ought to do *D*". For example, "Given that you have promised to do something, you ought to do it". An all-things-considered "ought" judgment would then have the form, "All things considered, *P* ought to do *D*" — in other words, "Given *C*, where *C* is all things considered, *P* ought to do *D*". This amounts to the claim that the prima-facie "ought" is basic and that the all-things-considered "ought" is to be defined in terms of the prima-facie "ought". We might then define the basic, prima-facie "ought" like this: "Given *C*, *P* ought to do *D*" means the same as "*C* gives *P* a reason to do *D*".

Exactly similar remarks can be made about the "ought" of expectation. There is a prima-facie "ought" of expectation and an all-things-considered "ought" of expectation. We say, "Going by the timetable, the train ought to be here in five minutes; but, given that the engineer is new at the job, the train ought to be somewhat later than that". We can define the basic form of the "ought" of expectation as follows: "Given *C*, it ought to be that *S*" means the same as "*C* is a reason to believe that *S*". Analogously, we might define the basic form of the evaluative "ought" as follows: "Given *C*, it ought to be that *S*" means the same as "*C* is a reason to wish or hope that *S*".

Finally, consider the "ought" of simple rationality. Kant holds that moral requirements derive from reason alone; according to him it is irrational not to act morally. If that were true, there would be no reason to distinguish the moral "ought" from the "ought" of simple rationality. And, even if we suppose that Kant is wrong about the powers of reason, we might still suppose that the moral "ought" is a special use of the "ought" of rationality — not that it is irrational to fail to act in accordance with the moral law but rather that moral "ought" judgments are judgments using the "ought" of rationality made about agents who are believed to intend to adhere to the relevant moral principles.



The basic logical form of an "ought" statement, using the "ought" of rationality, is "Given *C*, *P* ought to do *D*". The suggestion is that this is the moral "ought" if the conditions *C* include *P*'s intending to adhere to the relevant principles. In that case, in saying "Given *C*, *P* ought morally to do *D*" we are saying "Given both *C* and that *P* intends to adhere to the relevant principles, *P* ought to do *D*". According to this suggestion, the difference between the judgment that the bank robber ought to give up his trade and the judgment that the bank robber ought to use the rear door is that in the former judgment but not in the latter we take the relevant conditions *C* to include the bank robber's acceptance of certain principles. What we mean is that, given his acceptance of those principles, he has reasons to go home rather than to continue robbing the bank.

This is to reduce the moral "ought" of rationality. Analogous reductions of the other "oughts" can also be given. The "ought" of expectation in "Given *C*, it ought to be that *S*", becomes "Given *C*, one ought to believe that *S*". The evaluative "ought" in "Given *C*, it ought to be that *S*", becomes "Given *C*, one ought to hope or wish that *S*".

It still makes sense to say that the word "ought" has four different meanings, since these reductions are not the same from one case to the next. A sentence of the form "Given *C*, *P* ought to do *D*" might mean any of four different things, which can be expressed using the "ought" of rationality as follows:

"Given *C*, *P* ought to do *D*", "Given *C* and that *P* accepts the moral conventions we accept, *P* ought to do *D*", "Given *C*, one ought to believe that *P* does *D*", "Given *C*, one ought to hope or wish that *P* does *D*". These are not equivalent. Nor is it quite correct to say that the moral "ought" is just a special case of the "ought" of rationality. The difference is that when we use the "ought" of rationality to say that *P* ought to do *D*, we are not necessarily endorsing *P*'s doing *D*, but, when we use the moral "ought" to say that *P* ought to do *D*, we are normally endorsing *P*'s doing *D*. When I judge that the

bank robber ought to use the rear door, I do not endorse his doing so; I am not indicating that I am in favor of his doing so. But, when I say that the bank robber ought to give up his trade, I do endorse his doing that; I am indicating that I am in favor of his giving up his trade. The moral "ought" is therefore the "ought" of rationality plus something else. When I use the moral "ought" I presuppose that the agent and my audience accept certain practical principles that I also accept, and I make my judgment in relation to those principles. Consider also the ways in which we react on learning that an agent does not have the goals we supposed that he had. I judge that a bank robber ought to use the rear door because I suppose that his goal is to rob the bank and get away unobserved. If I learn that he does not intend to rob this bank but is merely making a deposit, then I withdraw my judgment that he ought to use the rear door and I say that I was mistaken. On the other hand, suppose that I use the moral "ought" to say that the bank robber ought to give up his trade. In saying this, I am presupposing that certain conventions that the bank robber accepts give him reasons to stop being a bank robber. If I learn that the bank robber is, however, totally amoral and that, given his goals and plans, he has absolutely no reason not to continue to be a bank robber, then although I will withdraw my judgment I will not do so by saying that I was *mistaken*. This difference indicates that the two sorts of judgment are not of the same kind. The moral sense of "ought" is distinct from the sense of "ought" of rationality, even though the two senses are closely related.

#### IV

It may seem, by the way, that I am oversimplifying when I speak of the *moral* "ought", since the same sense of the word is used when it is said that someone has reasons to do something in relation to rules of law, club rules, conventions of etiquette, rules of a game, and so forth, which the speaker takes the agent to accept. If, for example, it turns out that the

agent does not accept those rules or conventions, the speaker will not withdraw his original statement by saying that he was mistaken. I will, nevertheless, continue to speak of the moral "ought". The fact that "ought" has the same sense in all of these cases is additional support for the social convention theory of morality. Given that theory, it seems appropriate to say that those who accept rules of law, club rules, or conventions of etiquette accept them in the way that they accept moral conventions; indeed, I would say that they accept them *as* moral conventions, as part of their moralities. Similarly, I would say that people playing a game have adopted a temporary morality of something quite like it.

Another complication is that the moral "ought" can be used in relation to a morality that the speaker does not share and, in that case, in judging that *P* ought to do *D*, the speaker does not necessarily endorse *P*'s doing *D*. Consider such judgments as these; "You, as a Christian, ought to turn the other cheek; I, however, propose to strike back". A spy who has been found out by a friend might say, "I hope that you will not turn me in, although I realize that, as a loyal citizen, you ought to do so". In such a case, if it turns out that the agent is not a Christian, or is not a loyal citizen, the speaker can withdraw his original judgment by saying that he was mistaken. The difference between the moral "ought" and the "ought" of rationality does not emerge in such examples; it emerges only in the more usual case in which the speaker shares the relevant principles and endorses *P*'s doing *D*.

## RESUMEN

Muchas teorías éticas hacen depender la corrección de los juicios morales de la forma en la que esos juicios se relacionan con ciertos principios prácticos. Así lo hacen por ejemplo, las teorías de Kant, Hare y Sartre, así como la teoría de la convención.

De acuerdo con Hare, los principios prácticos pueden formularse como imperativos generales y los juicios morales correctos se siguen lógicamente del conjunto de estos imperativos generales en conjunción con asunciones fácticas verdaderas. Las teorías de Kant y Sartre y la de la convención difieren entre sí en cuanto a la naturaleza de los principios prácticos, pero cualquiera de esas teorías podría aceptar la idea de que los juicios morales tengan con los principios prácticos una relación deductiva.

Sin embargo, la idea de que deba haber este tipo de relaciones lógicas entre principios y juicios particulares no es muy convincente. En primer lugar, implica que los principios en cuestión incluyen suficiente detalle como para determinar lógicamente lo que tenemos que hacer en cada circunstancia concreta. Esta idea tiene un aire de irrealidad.

Por otra parte, esta teoría pierde de vista la variedad de juicios morales que existe. "*P debería hacer H*", "*P debe hacer H*", "*Sería bueno que P hiciera H*", "*Sería malo que P hiciera H*", etc. y no todos estos juicios son apropiados en las mismas circunstancias. La idea de la relación lógica parece más apropiada para juicios como "*P debe hacer H*" que a juicios como "*P debería hacer H*".

Esto sugiere una versión alternativa según la cual los principios prácticos en cuestión son vagos y carecen de la precisión legalista que sería necesaria para que pudiera darse la relación deductiva. Según esta teoría alternativa, más que principios que nos comprometen a ciertas acciones en ciertas circunstancias, lo que tenemos son propósitos y metas generales. Debemos respetar a los demás, debemos tratar de no hacerles daño, etc. . .

Aunque según estas teorías los principios son demasiado vagos como para que podamos deducir lo que debemos hacer (dadas ciertas circunstancias), de esto no se sigue que aceptar principios prácticos no nos provea de razones para actuar de cierta manera, la aceptación de principios prácticos puede darnos razones para actuar de tal o cual manera, de la misma forma en que lo hacen los demás propósitos, metas, deseos y planes que tenemos.

Según este punto de vista, entonces, decir que *P debería hacer H*

es decir que *P* tiene razones suficientes para hacer *H* que son más fuertes que las que tiene para hacer otras cosas. Si lo dicho es que *P* debería moralmente hacer *H*, lo que se afirma es que *P* debería hacer *H* por razones morales. Ahora bien, tener razones para hacer algo no es poder deducir ese acto de principios y circunstancias. La relación de tener razones es distinta a la relación de deducibilidad. Las razones pueden ser más o menos fuertes, y esto explica las diferencias entre distintos juicios morales, por ejemplo, entre “*P* debe hacer *H*” y “*P* debería hacer *H*”.

Las teorías de este tipo se han llamado “análisis de los juicios morales en términos de buenas razones”. Los juicios morales que se consideran son los que usan palabras como “debe” o “debería”. Según estas teorías, tales juicios dicen algo acerca de las razones morales que una persona tiene para actuar. Un análisis semejante no sería apropiado para otro tipo de juicios morales, tales como “*P* es malo moralmente en virtud de que hizo *H*”, porque estos juicios no dependen en la misma forma de asunciones acerca de las razones de *P*.<sup>1</sup>

Una ventaja del análisis en términos de buenas razones es que no sólo da cuenta de los usos morales de “debe”, “debería”, “puede”, etc., sino también del de otros usos como el “debe” de expectación, que se da, por ejemplo, en la oración “El tren debe haber llegado a las 5”. Este enunciado puede explicarse como afirmando que tenemos razones suficientes para creer que el tren llegó a las 5.

Otra ventaja del análisis en términos de buenas razones es que ayuda a explicar un aspecto de nuestro uso de “debería” que ha sido subrayado por W. D. Ross. Este autor dice que hay dos “deberías” morales: un “debería” prima facie, y un “debería” habiendo tomado todo en consideración. Al enunciar nuestros principios morales, dice Ross, no enunciamos principios absolutos: “Deberías cumplir tus promesas” no nos dice que siempre haya que cumplirlas: hay circunstancias bajo las cuales lo que deberías hacer es no cumplirlas.

La reacción de Hare ante este hecho es suponer que nuestros principios morales son en realidad mucho más complicados: todas las aparentes excepciones están ya consideradas. Ross, en cambio, nos dice que el principio moral usa el “debería” prima facie, nos dice que si hemos prometido hacer algo, esto nos da una razón moral para cumplir; si no hay razones opuestas, entonces deberías cumplir; si las hay, habrás de considerarlas todas y decidir lo que deberías hacer. Esto concuerda con un análisis en términos de buenas

<sup>1</sup> El lector podrá encontrar en *Diánoia*, 1975, la versión de Harman de “buenas razones” en *Una Teoría Naturalista de las Razones* (nota del resmisor).

razones. Lo que debemos hacer, considerados todos los casos, no está determinado por nuevos principios morales sino se decide en la forma en la que se sopesan razones en conflicto, una frente a otra, sean estas razones de orden moral o de cualquier otro.

Podríamos encontrar en la teoría de Ross una idea acerca de la forma lógica de los enunciados formulados con la palabra "debería". La forma lógica sería "Dado que *C*, *P* debería hacer *D*", lo que quiere decir "*C* le da a *P* una razón para hacer *D*". Por ejemplo "Dado que has prometido hacer *H*, deberías hacer *H*". El "debería" habiendo tomado todo en consideración puede definirse en términos del "debería" prima facie, basta interpretar *C* como todo lo tomado en consideración.

Se analizan en forma semejante los demás usos de "debería", tales como el de expectación y el de simple racionalidad. La diferencia entre "el asaltante de banco debería renunciar a este oficio" y "el asaltante de banco debería haber entrado por la puerta de atrás" es que en el primero las condiciones relevantes *C* incluyen la aceptación por parte del ladrón de ciertos principios. Así, el "debería" moral queda reducido al "debería" de la racionalidad. De igual manera se reducen el "debería" de expectación y el valorativo al "debería" de la racionalidad.

Pero puede seguirse sosteniendo que una oración de la forma "Dada *C*, *P* debería hacer *H*" puede querer decir cuatro cosas distintas en términos del "debería" de la racionalidad:

- (1) Dada *C*, *P* debería hacer *H*.
- (2) Dada *C* y que *P* acepta nuestras convenciones morales, *P* debería hacer *H*.
- (3) Dada *C*, deberíamos creer que *P* hará *H*.
- (4) Dada *C*, deberíamos esperar o desear que *P* haga *H*.

No debe pensarse que el "debería" moral es simplemente el "debería" de la racionalidad: es este debería y algo más. Presupongo que el agente y mis oyentes aceptan ciertos principios prácticos. Pero si descubro que el ladrón es amoral, y no comulga con nuestros principios prácticos, aunque retiro mi juicio no lo hago diciendo que estaba equivocado. Esto distingue este tipo de juicios de los de mera racionalidad. Si el agente no quería robar el banco, sino hacer un depósito, tengo que retirar mi juicio y tengo que admitir que estaba equivocado.

Parecería que el asunto se ha simplificado al hablar del "debería" moral porque el comportamiento del "debería" aplicado a la ley, a reglas de un club, a las convenciones de la etiqueta, a las reglas

de un juego, etc., es igual al del "debería" moral de acuerdo con lo que se ha dicho hasta aquí. Sin embargo, el hecho de que "debería" tenga en estos casos el mismo sentido que en la moral puede tomarse como una razón en favor de la teoría de la moral como convención social, más aún, podemos decir que aceptamos esas convenciones como convenciones morales.

Resumen por  
Hugo Margáin