MORAL OBLIGATION: FORM AND SUBSTANCE

Professor Stephen Darwall
Yale University

An advance draft of a paper to be delivered to the Aristotelian Society on
Monday November 9th 2009 at 4.15pm
Room G22/26, South Block, Senate House, Malet Street, London, WC1E 7HU

I believe it is possible to make an argument from the form that moral obligations
necessarily take to basic substantive propositions about what our moral obligations are. In
what follows I sketch how I think this can be done. By “moral obligation” I mean
something narrower than what moral reasons recommend, even conclusively, namely, what
morality requires or demands, in other words, what it would be morally wrong not to do. My
contention is that a proper understanding of what it is for morality to require or demand
something can enable us to derive at least some basic moral obligations.

I draw from an analysis of moral obligation that I defend in The Second-Person
Standpoint (Darwall, 2006). There I argue that moral obligation is conceptually tied to moral
responsibility, specifically, to a fundamental answerability we have to one another (and
ourselves) as representative persons or members of the moral community. Who are this ‘we’
and ‘us’? Anyone, I argue, with the psychic capacities necessary to enter into relations of
mutual accountability. I call these capacities second-personal competence, for reasons that will
emerge presently. Moral obligations are what we as second-personally competent persons
are answerable for to one another (and ourselves) as representative second-personally
competent persons. This, I shall say, is moral obligation’s form.

Because accountability to someone entails the latter’s authority to demand and bold
one to conduct (“second-personally”), I claim that accountability necessarily involves a
second-personal relation and authority. In the case of moral demands, the requisite standing is not any individual authority anyone has to exercise at his or her discretion. It is rather a representative authority we all, all second-personally competent persons, have to hold one another and ourselves to moral demands. We exercise this authority, I claim, when we have Strawsonian “reactive” attitudes like indignation or moral disapproval toward people and blame them when they violate moral obligations without excuse (Strawson 1968).

So far this says nothing about what we are morally obligated to do, about the substance of our obligations. Moral obligation’s form seems only to dictate that whatever its substance is is something we warrantedly expect of one another and ourselves. I shall argue, however, that form does indeed constrain moral obligations’ substance. But before I start, I should first set aside two arguments to this conclusion that I reject.

Some philosophers have held, first, that other-regarding content of some kind is simply built into the concept of morality or moral obligation. Whether the idea is that morality essentially has the point or “object” of ameliorating human or other sentient life or of solving collective action problems that arise from the unrestrained pursuit of self-interest, views of this kind make “teleological” content intrinsic to the concept of morality (Foot 1958, Warnock 1971, Baier 1958; see also Falk 1965). My argument will depend on no such alleged teleological conceptual truths, which I am inclined to reject, in any case. I shall assume that the concepts of morality and moral obligations are best identified formally, by their role in our attitudes and practical reasoning, not by any beneficial point or purpose.

Second, it might be thought that by analyzing moral obligation in terms of accountability, I have already conceptualized it in terms of what we owe to one another and so implicitly assumed that moral obligations are in their nature “directed” or “bipolar,” thereby constraining any content or substance they might plausibly be supposed to have to familiar
obligations we have to each other (Scanlon 1998, Gilbert 2006, Thompson 2004). It is important to see, however, why this is not the case. Seeing this, moreover, will also help us focus more precisely on the central issue with which we shall be concerned.

As Hohfeld pointed out, any (bipolar) obligation that A owes to B entails a correlative claim right that B has against A (Hohfeld 1923: 65-75). I follow Feinberg (1985) in holding that claim rights are to be understood in terms of a distinctive standing or authority that the right holder (or obligee) has to make claims against the person obligated to her (the obligor) and, I add, to hold the obligor accountable to her for complying with them. This means that bipolar obligations to someone are to be understood in terms of an individual authority that the obligee or right holder uniquely has to, among other things, claim her right or forego it by consent, hold her obligor answerable and resent the injury if her right is violated, seek remedy or not at her discretion, and, perhaps, to forgive. Only she or her representatives can do these things.¹

Now if, as I said at the outset, the form of moral obligation period, as we might call it, involves a fundamental answerability we have to everyone, ourselves included, as representative persons, hence their and our representative authority, then bipolar obligations to others, even, any we might have to every other person, implicate a different standing or authority that others have as individuals to hold us answerable, hence their individual authority. Assuming that the form of moral obligations period involves a fundamental answerability to one another and ourselves as representative persons, therefore, does not entail that they are, or even that they entail any obligations that are, owed to one another. Bipolar obligations to others implicate individual authority, whereas moral obligations (period) involve representative authority.

¹ It is a reflection of this that it is up to the injured party whether or not to bring a tort action at law, whereas criminal actions are appropriately brought by the people and their representatives.
Still, the reverse cannot be true; individual authority must surely entail representative authority. If A is genuinely obligated to B to do X, then A must be also be morally obligated to do X period, at least other things being equal. To put it in other words, it seems that an action cannot violate a right or bipolar obligation and hence *wrong the obligee*, unless it would be *wrong* period, at least other things being equal.² So if an action wrongs someone, it violates both the victim’s individual authority as well as a representative authority she shares with the wrongdoer.

We are concerned, however, with the relation in the other direction. We have seen that nothing in the idea moral obligations as involving answerability to one another as representative persons entails anyone’s individual authority of the kind that is conceptually implicated in bipolar obligations to and correlative claim rights against. For example, although a moral obligation not to foul the environment involves, on the analysis I am proposing, our being answerable to one another for avoiding environmental degradation, that does not necessarily mean that this is something we *owe to each other*, since no one may have the individual authority to demand this conduct or not at his individual discretion. Moreover, as far as the concepts go, it might be that *all* moral obligations are obligations (period), there being no valid obligations to or rights against. Although individual authority apparently entails representative authority, the converse does not seem to hold.

**MORAL OBLIGATION AND ACCOUNTABILITY**

I believe that an argument can nonetheless be made from moral obligation’s form as fundamental answerability to one another as representative persons to a basic substantive moral obligation period, *and* to reciprocal, bipolar obligations, to respect and not undermine one another’s moral autonomy. Before I begin to make the argument, however, I need to

---

² I say “other things being equal,” because not all wrongings, hence violations of bipolar obligations and correlative claim rights involve wrongdoing, all things considered. On this point, see Coleman (1992: Ch. 11).
say more about the general outlines of the account of moral obligation from which I shall be drawing.

Formal accounts have tended to be advanced in terms of moral obligation's categorical character and/or its putatively supreme authority. While both features seem necessary, we can fairly easily see that they are, even jointly, insufficient.

Note first that a conceptual gap exists between the ideas of, on the one hand, what morality recommends and, on the other, what it demands or requires. To see this, ask yourself whether you think that the possibility of supererogation is conceptually open. In other words, would someone who thinks that an action might be conclusively morally recommended but not morally required or not wrong not to do, be involved in some kind of self-contradiction or conceptual incoherence? It seems clear she would not.

Whether supererogation is possible seems clearly to be a substantive normative issue that is not settled by the concept of moral obligation alone. But if that is so, then moral obligation must be a distinct concept from that of what morality recommends, even conclusively. But notice next that moral reasons are no less categorical when they recommend conduct than when they demand it—their validity and recommending power is no less independent of their relation to the agent's ends. And finally, notice that, so far as the concepts go, moral reasons might recommend an act and also outweigh reasons for acting otherwise without its necessarily being the case that the action is morally required or that not doing it is morally wrong. It seems conceptually possible that there might be cases where an agent does nothing wrong though she acts against the balance of moral reasons, even when these outweigh reasons of other kinds. Such an agent would of course be subject to rational criticism, but not necessarily, it seems to moral blame.
If all this is correct, then the concept of moral obligation must differ from that of what moral reasons recommend, however categorical or overriding these reasons might be. What fills the gap, as I see it, is the concept of accountability and moral blame. Moral obligations are moral demands, and moral demands are whatever we are legitimately held to or answerable for, that is, where violations are blameworthy if the agent lacks a valid excuse. As Mill put it, “we do not call an action wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it” (Mill 1998: Ch. 5). “Punishment” may sound overly strong, but Mill includes blame and “the reproaches of [the agent’s] own conscience.” What it is, consequently, for an action to be morally obligatory and its omission to be morally wrong, is for it to be something whose omission would warrant blame and feelings of guilt, were the agent to omit the action without excuse.³

ACCOUNTABILITY, AUTHORITY, AND SECOND-PERSONAL DEMANDS

I will assume that an analysis of moral obligation that ties it conceptually to accountability and warrant for reactive attitudes like blame and guilt is broadly correct. Our question will then be whether anything about what we are obligated to do follows from this account of its form. I shall argue that it does, more specifically, that an argument can be made from the equal representative authority implicit in this account of moral obligation’s form to a fundamental obligation to respect and not undermine moral autonomy as well as to a basic reciprocal, bipolar obligation that entails every person’s individual authority to demand such treatment.

I begin with some points about moral responsibility that derive from P. F. Strawson’s “Freedom and Resentment” (Strawson 1968). Strawson coined the term “reactive attitude” to refer to a set of mental states that play a distinctive role in holding

³ So moral wrongness and blameworthiness are distinct concepts, since an action may be wrong though not blameworthy because of the existence of an excuse.
people morally responsible, whether another person, as in indignation, resentment, or moral blame, or oneself, as in the emotion of guilt. Strawson didn’t give a formal definition of these attitudes, but their central features are clear from their role in his argument about moral responsibility and freedom of the will. Strawson’s central idea is that reactive attitudes involve a way of regarding the individuals who are their objects that commits the holder of the attitude to certain assumptions about the object individual and her capacities to regulate her will. Unlike “objective attitudes,” like disdain, disgust, and annoyance, reactive attitudes are essentially characterized by “involvement or participation with others in inter-personal human relationships” (Strawson 1968). There is always an essentially “interpersonal,” or as I shall say, “second-personal” element to reactive attitudes. Through the attitude we hold its object to something and thereby implicitly make a demand of him or her. As Strawson puts it, “the making of the demand is the proneness to such attitudes” (Strawson 1968). The reason that reactive attitudes distinctively implicate freedom of the will, then, is that we can intelligibly address a demand to someone to regulate her will appropriately only if we suppose that she can so regulate it as a result of recognizing our demand’s legitimacy. The supposition is a “constraint of moral address” (Watson 1987: 263,264).

Consider the difference between the kind of disdain expressed by the put down, “He is ‘an egregious blockhead’” (in Hume’s memorable phrase (Hume 1985)), and an attitude of indignation or moral blame (as might intelligibly be expressed even by, “Stop acting like an egregious blockhead”). Unlike the latter, the former is unfettered by any constraints of address that inevitably arise when we take a second-person perspective toward someone. Disdain is not standardly addressed to its object; if it, indeed, has any addressee, it is more likely to be others we think capable of appreciating why its object is a worthy target. When we blame someone for something, however, we implicitly make a demand of him to act
differently and, if he has not, to take responsibility for not having done so, where taking
responsibility is essentially an interpersonal matter (holding himself answerable to us (and,
indeed, to himself) as representative persons). In so regarding him, we perforce see him as
intelligibly so regarded, as someone, who is competent to take such an attitude toward himself
and guide himself by it. We see him as capable of entering into reciprocal human
relationships of mutual accountability.

Disdain involves no such assumptions. One can hardly imagine Hume, believing
that his disdain’s object is too thick to appreciate his own blockheadedness or that he is
incapable of changing it, withdrawing his put down as not expressing a fully intelligible
attitude. Moral blame, on the other hand, holds its object to a demanded standard and to its
object’s holding himself to that standard by making himself answerable for compliance. So
it is not fully intelligible, or, at least, it is unwarranted in its own terms and not just unfair,
when its object is someone we know lacks the psychic capacities necessary to do this. It just
doesn’t make sense to blame someone for dull-wittedness, unless one is under some illusion
about the human ability to take responsibility for mental endowments. No matter how
regrettable or unfair, however, disdain for dull-wittedness is surely an intelligible attitude.

Strawson makes an important distinction within reactive attitudes between “personal”
and “impersonal” ones. This can be confusing, since it is possible to lose track of the fact
that all reactive attitudes, even impersonal ones, are “inter-personal” (or “second- personal”).
“Personal” reactive attitudes are those, like resentment and guilt, that are felt as if from the
perspective of a participant in the events giving rise to it, whereas “impersonal” reactive
attitudes, like indignation or moral blame, are felt as if from a “third party’s” point of view.
One cannot resent or forgive injuries to people with whom one lacks some personal

---

4 Guilt, as will become clear presently, is interpersonal (or second-personal) in this sense. See note 6.
connection, but this is no impediment to moral blame or disapproval. Nonetheless, however “impersonal,” blame is not an “objective” attitude in Strawson’s sense. It is just as “inter-
personal” or “second-personal” as personal reactive attitudes like resentment or guilt. Thus although impersonal reactive attitudes are as if from the perspective of a “third-party,” they are not “third-personal” attitudes in the usual sense; they involve the same second-personal element of implicit address as do personal ones, only as if from the perspective of a representative person rather than any individual’s standpoint.

The difference between warranted personal and impersonal reactive attitudes tracks the distinction between individual and representative authority and, consequently, between bipolar obligations to and moral obligations period. On the account we have been assuming, it is a conceptual truth that unexcused violations of moral obligations period warrant impersonal reactive attitudes like indignation and blame. Likewise, it is a conceptual truth that unexcused violations of bipolar obligations to someone, and hence of the latter’s claim rights, warrant personal reactive attitudes like resentment from the obligee or right holder, at least other things being equal.

The same points can be put by saying that personal and impersonal reactive attitudes implicate different authorities to make the demands they implicitly address. When we feel moral disapproval or blame towards people for violating a moral obligation (period), we implicitly address demands not as individuals, but as representative persons; indeed we implicitly demand that those we blame make the same demand of themselves from the very same perspective. When, however, someone has violated a claim right you had against him,

---

5 “The same abnormal light which shows the agent to us as one in respect of whom the personal attitudes, the personal demand, are to be suspended, shows him to us also as one in respect of whom the impersonal attitudes, the generalized demand, are to be suspended” (Strawson 1968)

6 Similarly, second personal does not imply second party. Guilt, like any reactive attitude is second personal, since it involves implicit address, but it clearly is not a second party attitude. In feeling guilt, one implicitly addresses a demand to oneself. Finally, any second-personal attitude is also first personal. Address, whether implicit or explicit is always from someone (an individual (I) or a collective (we)).
hence an obligation he had to you, he has not just done wrong—he has wronged you. So you have a distinctive individual standing to hold him answerable as the victim, for example, to resent the wrong, which you can exercise or not at your discretion. You can seek compensation or not, forgive or not, and so on. No one else has the standing to do these things, except perhaps to represent you if you cannot speak for yourself. To sum up:

*Warranted impersonal reactive attitudes, and thus moral obligation period, implicitly involve representative authority; and warranted personal reactive attitudes, and thus obligations to and correlative claim rights against, implicitly involve individual authority.*

**SECOND-PERSON COMPETENCE AS SUFFICIENT FOR MORAL OBLIGATION**

We are now in a position to begin a positive argument. We can begin with Strawson’s point that when we take someone to be morally obligated and hold him responsible with an impersonal reactive attitude, we have to attribute to him whatever psychic competences are necessary to enter into mutually accountable, interpersonal relationship.⁷

What, then, does second-personal competence involve? In addition to whatever capacities we might include under “rational agency,” I take it to follow from Strawson’s argument that, for human beings, at least, it includes the capacity for reactive attitudes, in addition to imaginative and empathetic abilities necessary to put oneself in others’ shoes, including those of a representative person. But further, second-personal competence also includes the capacities to make and regulate oneself by normative judgments about what attitudes are warranted from these points of view. This need not mean that moral obligation

---

⁷ This is often not an all or nothing affair, of course. But we can abstract from these complexities and assume that what we might call “full” moral obligation and moral responsibility presuppose “full” second-personal competence (“second-personal competence” for short). So understood, second-personal competence is what Rawls called a “range property” or concept, much as Rawls treated the capacity for “moral personality” in *A Theory of Justice* (Rawls 1971: 509). All that matters is that one is above a certain threshold.
is tied conceptually to human psychology. So long as a being has the capacity to take a second-person standpoint toward others and himself, make judgments about what demands would be warranted from this perspective, and regulate his conduct through making the relevant demands of himself, the being counts as second-personally competent.

The Strawsonian insight, as I interpret it, is that second-personal competence is both necessary and sufficient for moral obligation. Strawson clearly holds that second-personal competence is necessary. But the thought that it is also sufficient finds no less support in the reflections he relies on in linking moral obligation and answerability.

It is uncontroversial that fundamental moral obligations apply equally to all moral agents. When utilitarians say that we are morally obligated to do what will maximize utility or to comply with rules whose acceptance maximizes utility, they do not mean that we are so obligated as human beings, as members of a certain biological species. Like other moral theorists, they apparently take it that there exists a set of capacities, which we collect under the term “moral agency,” that are both necessary and sufficient to make one subject to moral obligations. Not every human being has these capacities, indeed, though it is uncontroversial that most adult human beings do. Nothing ties moral agency to species membership.

It follows from the letter of Strawson’s argument that moral agency includes second-personal competence. Otherwise second-personal competence wouldn’t be necessary for moral obligation, since moral agency, however we understand it, is supposed to be sufficient. The only way that second-personal competence might then not be sufficient for moral obligation is if moral agency included additional features, features that, when added to second-personal competence, make up an ensemble of capacities sufficient to make whomever has them subject to moral obligations.
What might such additional features be? No specific knowledge seems to be explicitly required. We do think, of course, that people can be held moral responsible, and so be properly held subject to moral obligation, only on the assumption that they can know, appreciate, and regulate themselves by the obligations with which we take them to be answerable for complying. But we do not take it as a defense or excuse that someone simply lacked the relevant knowledge, at least when it concerns the moral obligations of the most basic sort. We presume that moral agency includes the capacities necessary to acquire and appreciate this knowledge, at least in principle.

Neither do we suppose that any specific contentful motivational state, like benevolent concern, is necessary to be subject to moral obligation. Our sense that we can legitimately demand that people not harm others, for example, is in no way conditional on agents being able to care about their potential victims and their welfare for their own sakes. We do not treat lack of benevolent concern as an excuse or mitigating factor. Even beings who, like Kant’s “unfortunate man,” who are “by temperament cold and indifferent to the sufferings of others” would seem to have some basis to appreciate the legitimacy of the moral demand not to harm, for example, if they would make such a demand of others and can see that no such demand can be warranted except impartially (Kant 1998: 438).  

If moral agency involves no additional features, however, and if second-personal competence is necessary for moral responsibility and obligation, it follows that the capacity for moral agency is nothing other than second-personal competence. The latter is both necessary and sufficient for moral obligation.

Somewhat unexpected support for this thought comes from Hobbes. After listing the laws of nature in Leviathan, Hobbes writes:

---

8 I do not mean, however, that forms of psychopathology that make one dead to moral considerations or to impartially regulated reactive attitudes cannot excuse.
to leave all men inexcusable, they [the laws of nature] have been contracted into one easy sum, intelligible, even to the meanest capacity; and that is, Do not that to another, which thou wouldst not have done to thyself (Hobbes 1994: 15.¶35).

Hobbes must be thinking, along the Strawsonian lines we have been pursuing, that duties imposed by laws of nature are essentially tied to responsibility. If they could not be “sum[med]” into something “intelligible, even to the meanest capacity;” violations would be excusable quite generally. But notice also the form that Hobbes’s summation takes. Hobbes cannot plausibly mean by “wouldst not have done to thyself” what one would prefer not to have done to oneself. I may prefer a world in which others peel me grapes, and, if so, there is a sense in which I would not “have” them forbear to do this. But this would give me no reason to think that I have any duty to peel grapes for others.

Suppose, however, that I expect others to peel me grapes, hold them to this demand, feel indignant and resent it when they do not, and so on. This must be the sense that Hobbes has in mind. The sum of the laws of nature is that one should not act towards others as one would expect or demand of others that they not act toward one and would blame them for doing, were they so to act without excuse. Knowledge of our fundamental moral obligations must apparently take some such form if second-personal competence is to be sufficient for it.

**SECOND-PERSONAL COMPETENCE AS GROUND FOR MORAL OBLIGATION**

An obvious problem with formulations of the golden rule as a fundamental moral principle that has long been noticed is that so long as people are prepared to accept others making the same demands of them, nothing rules out their making even the most outrageous demands of others. As Hobbes formulates it, the golden rule is less a fundamental axiom from which basic moral obligations can be derived than a principle of moral consistency or
integrity. Still, it highlights moral obligation’s second-personal character. Action toward others can be consistent with our moral obligations as we understand them only if it is consistent with demands we would make on others (and, of course, we make only demands we take to be warranted).

We have been assuming that what matters for moral obligation is what demands would be warranted from the impartial perspective we and others share, namely, that of representative person. Moreover, if others are bound by the very same fundamental moral obligations we are, then others must be able to see these demands as warranted from the very same perspective also. It follows that something can be a fundamental moral obligation, then, if, and only if, it involves a demand of anyone (i.e., any second-personally competent person) that anyone can see to be warranted from the perspective a representative second-personally competent person.

For moral obligations to exist at all, consequently, there must be fundamental demands that anyone can see to be warranted from anyone, as a representative second-personally competent, to any second-personal competent, himself or others. But for this to be true, there must be something that could be seen by anyone subject to them to warrant such demands. What could provide such warrant?

Now the process of considering what demands are warranted from the perspective of a representative person is itself an exercise of second-personal competence. So when we hold one another and ourselves responsible from this perspective through exercising second-personal competence, there is a straightforward sense in which we are committed to valuing one another’s exercise of this very same competence. A central lesson of Strawson’s discussion is that answerability is always a reciprocal, second-personal affair. It is impossible warrantedly to hold others answerable without allowing them the interpersonal space to take
responsibility for themselves by making their own autonomous choices. One cannot coherently hold someone responsible for (freely choosing) to do something while simultaneously undermining his ability to make the choice himself or otherwise usurping his agency.9

The interest in living with others on terms of mutual accountability, exercising and allowing them the liberty to exercise the ability autonomously to make their own moral choices and so take responsibility for themselves as their author is thus built into the second-person perspective of a representative person. Since this is so, and since fundamental moral obligations are demands from that perspective, there must then be a fundamental moral obligation not to usurp others’ moral agency or otherwise undermine the conditions of free moral choice. Here already we have derived some substance from moral obligation’s form.

And we can go further. It simply cannot be the case that there is only this one moral obligation, since there can be significant moral choices only if there are others also. But here again, any other moral obligations would have likewise to be things any second-personally competent agent can find warrant for from the perspective of a representative person, through exercising second-personal competence. It seems inconceivable, from this perspective, that the interest in autonomous choice can be contained within some specified moral realm. Part of what it is to respect someone’s standing to make her own moral choices, even as one makes moral demands of her, is to leave it up to her how and whether to comply. Otherwise, she will not be complying of her own free choice and so holding herself responsible. This means that the interest in exercising second-personal competence involves an interest in being accorded the interpersonal space necessary to lead one’s own life more generally. And if that is so, it follows that pro tanto demands not to violate

9 For an interesting analysis of paternalism as involving illegitimate usurpation of agency, see Shiffrin 2000.
autonomy and restrict liberty are also warranted more generally. And there is no reason to think that warranted demands grounded in this interest would not also extend significantly wider, to include, for example, obligations of mutual aid and non-maleficence, as well as others.

Moreover, this interest can also ground an individual authority to lead one’s own life of the sort involved in a claim right, that is, not just the wrongness of usurping moral autonomy or otherwise violating the conditions of free moral choice, but also the wrongness to the individual moral agents involved. There is absolutely nowhere to stand to object reasonably to any second-personally competent agent’s claim to make his or her own moral choices. We can hardly warrantedly hold him responsible as representative persons for taking individual responsibility for himself without thereby committing ourselves to according him a distinctive individual authority to hold us responsible for letting him do that. Thus although we can distinguish analytically between a representative person’s perspective from which we and he have the very authority for holding people responsible for letting people make their own moral choices and his individual standing to object to our failing to let him do so, it would seem that any interest in the exercise of autonomous moral choice that can provide warrant for the former impersonal reactive attitudes and demand must also be able, impartially, to warrant the latter personal reactive attitudes and demand as well.

Obviously, this is only a beginning sketch of how one might seek to derive moral obligations and claim rights from materials that are available within a second-person standpoint. I hope it is clear, however, that if moral obligation’s form consists in a fundamental answerability we have to one another, then its substance must include basic obligations to allow one another to take responsibility for ourselves.

Stephen Darwall
Yale University


