

The Moral Responsibility of Firms

Edited by
ERIC W. ORTS
and
N. CRAIG SMITH

OXFORD
UNIVERSITY PRESS

The Conversable, Responsible Corporation

Philip Pettit

This chapter offers a defense of a simple syllogism. First premise: corporations are conversable agents. Second premise: conversable agents are fit to be held responsible for their actions. Conclusion: corporations are fit to be held responsible for their actions. The argument will proceed briskly, since I have defended some of the elements it deploys in previous work, much of it done in collaboration with Christian List.¹ The chapter is in three sections, two devoted to the premises, the third to the conclusion.

FIRST PREMISE: CORPORATIONS ARE CONVERSABLE AGENTS

What Are Corporations?

Corporations—or firms or companies or businesses, as we may also call them—are a species of corporate organization, on a par with clubs, churches, universities, voluntary associations, political parties, and indeed political states. Like such other organizations, as we shall see, they count as agents; that is what distinguishes them from informal groups. But unlike the bulk of other corporate bodies they are commercial, profit-seeking organizations (Orts 2015). And unlike other commercial organizations such as partnerships, they are distinguished mainly by asset lock-in (Kraakman 2009; Ciepley 2013). The members who provide the finance that any corporation will require—the

¹ See (Pettit 2003; 2007; 2009; List and Pettit 2011; 2012; Pettit 2014). Particularly important from the viewpoint of this essay is (Pettit 2007), although the line I pursue here is different in character.

shareholders—cannot remove their support at will. They can sell their shares to others, of course, if they can find buyers. But they do not have even the qualified ability to take their capital out of the corporation and use it for their own purposes.

Asset lock-in is of the greatest importance in the organization of a business, since it means that a corporation can rely on retaining the capital at its disposal, so that it is not vulnerable in the manner of the normal partnership. But two other features are also characteristic of the corporation, if not exclusive to it. One is limited liability and the other entity shielding. Limited liability means that the individual shareholders of the company are not personally liable—their personal resources are not at risk—if the company declares bankruptcy. And, complementary to this, entity shielding means that the company is not corporately liable—its corporate resources are not at risk—if individual shareholders declare bankruptcy.

These features of a corporation have the effect of juridifying it. They force us to speak of it as a legal entity separate from the different categories into which members now divide: as an entity that its directors govern, that its shareholders invest in, and that employs its managers and workers. This makes for a contrast with some other corporate bodies. Where we may equate a regular corporate body with the membership as a whole—we may say, for example, that what a club does the membership does—there is no class of members with which we can identify the corporation in the same way; we have little option but to speak about it as an entity apart.

It is worth noting these distinctive aspects of corporations in order to place them in relation to other corporate bodies and to register their special character. But I note them here only to set them aside. The points defended in this chapter do not depend on anything that is exclusive to corporations as such and go through for any kind of corporate body. While my focus is on corporations, then, I shall often indicate that the claims I defend apply to corporate bodies more generally.

The Notion of Agency

The first thing to notice about corporations is that like other corporate bodies, and unlike unincorporated groups, they count in a straightforward sense as agents. But what is an agent? The observation that corporations are agents means nothing in the absence of an answer to this more general question.

In order to answer the question, think about an imaginary, simple artifice and the considerations that might lead us to describe it as an agent (List and Pettit 2011, ch. 1). Imagine a little robot that stands six inches high, has bug-like eyes protruding on stems, can move about on wheels and lift things with little arms at its sides. Suppose that, put on a table top strewn with half a dozen

bottles, this object seems to scan its surrounds, with its eyes turning now in this direction, now in that, and responds immediately to a particular stimulus: the appearance, as we take it to be, of a bottle on its side. The stimulus leads it to move to the bottle and, using its arms, to put the bottle upright. Or at least it does this when the lights are on; when there is a real bottle there, and not just the photograph of a bottle; and when the bottle is not so near the edge that the lifting procedure knocks it off the table.

We naturally hold that this robot is an agent—this, in a more or less common sense of the term—because of three things that are true of it. First, we can reasonably ascribe representations to it of its environment, since it reliably reacts over different sorts of stimuli to the appearance of a bottle on its side—even the misleading appearance, as the photograph shows—and does nothing when the bottles are, and presumably appear to be, upright. Second, we can reasonably take it to have the purpose of putting the bottles in its space upright, since it takes steps to make this the case when its representations indicate that some bottles are on their sides and leaves things be when its representations indicate that none are in that position. And third, it more or less reliably acts to realize its purposes—its single purpose, in this case—according to its representations. It forms suitable representations and it performs suitable actions in realizing its purposes, at least when conditions are normal by independently plausible criteria. It occasionally fails to operate at par but only when things are clearly rigged against it: the lights are off, there is a photograph placed to mislead it, or the bottle is near the edge of the table.

Generalizing from this example, we may say that a system is an agent when, in normal, unrigged circumstances, it maintains certain purposes, forms reliable representations of its environment, and acts reliably so as to satisfy its purposes according to those representations. In order to be an agent, then, the system must be evidentially reliable in the formation of representations—this, in the way the robot reliably notices bottles on their sides—and executively reliable in the performance of actions that, according to its representations, are appropriate ways of realizing its purposes. It must exemplify a unified epistemic vision and a coherent practical agenda (Rovane 1997).

In this sense of agency, it should be clear that while trees and plants do not count as agents, non-human animals like birds and cats and dogs certainly do. They have many more purposes than our little robot and many more ways of forming representations: they have many sensory channels, after all, and are capable of integrating the information these convey. But still, they can be said in the same way as the robot to form representations of how things are and to pursue their purposes according to those representations. Or at least they can be said to display such evidential and executive reliability—if you like, such evidential and executive rationality—in presumptively normal conditions.

Corporations Are Agents

Corporations, like other corporate bodies, count as agents in this generous sense. They may fail in the manner of any agents, when things are not normal by plausible criteria. But absent such distorting conditions we can identify various purposes by which they hold, perhaps ordered in a rough hierarchy of importance; we can take them to form reliable judgments or representations about the opportunities at hand or the best means of realizing those purposes; and we find that they more or less reliably act so as to satisfy their purposes according to their representations. Unlike animals, of course, corporations are not biological organisms and unlike robots, they are not electronic systems; they operate on the basis of psychological adjustments among their members, not on the basis of any natural or mechanical infrastructure. But like those other entities they fit the functional profile of agents, and that is all that matters.

The evidential reliability of the corporate body gives us a basis for predicting how its judgments would respond to new evidence, and its executive reliability a basis for predicting how it would act in the event of such a response. In decision-theory terms the organization has a kinematics that governs how it would update in such counterfactual situations, with individual members making the psychological adjustments that support appropriate responses.² This kinematics is grounded in a central, decision-making structure, as Peter French (1979) calls it, albeit a structure that may operate via a widely distributed network of sub-structures (Hess 2014).

These observations argue strongly for the view that corporations and other corporate bodies ought to count as agents. Those who reject this view tend to invoke a parallel between markets and corporations, arguing that they each amount to nothing more than a nexus of contracts, as it is routinely put, and that a nexus of contracts cannot exercise agency. The idea, then, is that just as we speak figuratively in saying that markets think and do things—say, think that banking stocks are overvalued and act to put a correction in place—so we speak only figuratively in ascribing attitudes and actions to corporations. On this approach, as one commentator puts it (Grantham 1998, 579), a corporation is just “a collective noun for the web of contracts that link the various participants.”

The appeal to the analogy with markets overlooks the presence of a decision-making structure in a corporation and the support this gives to a kinematics. Sure, that structure is grounded in the contractual relations between members but that does not argue in any way against the agential status of the corporate body. Those who invoke the analogy with markets to

² A system may have a more or less robust kinematics, depending on the range of counterfactual situations where it behaves appropriately. Agency requires an intuitively robust kinematics but it is hard to identify any threshold that a system must reach in order to count as an agent.

deny that status to corporations appear to be moved by the following invalid inference: corporations and markets are both composed out of contracts; but markets are not agents; so corporations are not agents. That argument, of course, is no more persuasive than this blatant non sequitur: animals and trees are both composed out of cells; but trees are not agents; so animals are not agents.³

To argue along the lines we have been following so far is to maintain that corporations are interpretable—and, by standard methodological criteria, best interpreted—as having certain purposes, forming certain representations, and acting so as to satisfy those purposes according to those representations. It is to say that like animals and robots they are best seen from within the intentional stance, as Daniel Dennett (1987) calls it (Clark 1994; Tollefsen 2002). But this is not yet to say all that needs to be said about corporations as agents.

Whereas animals and robots are mute beings, corporations are able to speak for themselves and to offer an interpretation of their own purposes, representations, and actions. They are not just interpretable as agents; they interpret themselves as agents. And they do not just interpret themselves as agents; they make a point of communicating that interpretation to their members and to those who interact with them in any way. Corporations speak for themselves in this manner when they rely on authorized spokespersons to speak in their name; these will be individuals or committees recognized under the organization's rules as having the authority, subject to some overall coordination, to express the corporate view in this or that domain. Such spokespersons will give an account of the corporation's goals, long-term and short-term; offer a narrative and usually a justification of its judgments about current market conditions and likely developments; and make sense of what the corporation does in the light of those imputed purposes and representations. And, in specific interactions with other agents—its members or employees, its customers or suppliers, the local government or tax authority—they will declare its intentions, formed on the basis of agreed assumptions, to perform in this or that fashion; in short, they will commit the corporation in a contractual manner.

In speaking for themselves in this way, corporations resemble individual human beings like you and me. And the similarity is important because it means that like you and me, corporations can use words as the means of forming their purposes and representations, not just as signs to report attitudes that are already formed.⁴ It means, in a word, that they are *conversable agents*.

³ I cannot consider other, more substantive arguments against the idea that corporations are agents. For two such arguments see, for example, Velasquez (1983) and Rönnegard (2013).

⁴ As will become apparent as my argument unfolds in the text, corporations typically form intentional attitudes on the basis of members subscribing to suitable words. Unlike human beings, and of course unlike non-human animals, they do form few intentional attitudes in

The Notion of Conversable Agency

When I ask you about what you believe or desire or intend to do you may speak about yourself as you would about a third person, speculating about precisely what attitude you hold and looking at your past behavior and current intentions in order to help you out in the exercise. But that is not what you or anyone else does in the normal case. What you ordinarily do on being asked whether you believe that such and such is to think about whether you ought to believe such and such—whether the data support the claim that such and such—answering that you believe it, disbelieve it, or have no opinion either way, depending on the evidence at hand (Anscombe 1957; Evans 1982). And equally, on being asked whether you desire or intend to do so and so, what you ordinarily do is to think about whether the prospect is one that you ought to take up in light of the desiderata that weigh with you—your concerns or interests—and, depending on your judgment, answer that you do or do not desire or intend it.

You may have had the relevant belief or intention prior to my asking you about it, of course. But the normal response to the query will still be to check before responding on whether you ought to have it and to form the attitude or to re-form or re-confirm it only in the case of a positive answer. In such a response, then, you do not give a report on your attitude as you might give a report about someone else's. You avow the attitude, as we might say, conveying it in a way that marks your capacity to form, re-form, or indeed unform what you believe and desire, as the relevant data or desiderata require. You speak on the basis of a maker's knowledge of your attitudes, as it used to be said, not a reporter's.

The striking difference between reporting an attitude and avowing it in this sense shows up in the event of your not acting as that attitude would have required or supported. In the case of a reporting a certain belief or intention, or as you might report the belief or intention of another, you can explain and excuse your failure to act appropriately in either of two salient ways. First, by arguing that you changed your mind between the time of reporting and acting; in this case, you may be expected to justify the change of mind by reference to new evidence or other considerations that came to light. Or, second, by arguing that you must have been misled by the evidence about what you believed or intended when you gave your report: this, in the way in which you might have been misled by the evidence about someone else's belief or intention.

Things are very different if you avow the belief or intention rather than reporting it. You might do this by affirming "p" in response to a question as to

a sub-personal manner: the only exceptions are those attitudes, like a belief in *modus ponens*, that are ascribed on the basis of their procedures.

whether you believe that *p*, or by asserting "I'll be there" in response to the question as to whether you intend to go to a party this evening; in each case the utterance signals that you have made up your mind, not merely that you found you were minded in this or that manner. In the case of such an avowal, you may excuse a failure to act appropriately by the changed-mind excuse but you will not be able to do so plausibly by appeal to the misleading-mind excuse. Because you signal in making an avowal that you are making up your mind rather than just conducting a mind-reading exercise, you cannot plausibly excuse a failure to live up to the words you utter by claiming that you must have gotten your mind wrong; that excuse would only have been available if you had presented yourself as someone reading your mind in the way in which you might try to read the mind of another.

We have good reason as individuals to avow beliefs and desires and intentions rather than just reporting them. By doing so we make our words in expression of those attitudes more expensive; we increase our exposure to a penalty for failure, manifestly depriving ourselves of an excuse whereby we might get off the hook. And by making them more expensive we make them more credible, thereby giving ourselves a better chance of persuading others to rely on us and to enter relationships of cooperation. In putting aside the misleading-mind excuse, we may be said to commit ourselves to the attitudes we communicate; we back ourselves, on pain of penalty, to display those attitudes. And we can also go one stage further in the case of intentions and, putting aside the changed-mind excuse as well as the misleading-mind excuse, commit ourselves in the full-scale manner of making a promise to act in a certain way. This extra commissive move will have the attraction of making our words even more expensive and more credible.

Serving us well in the formation of relationships, self-commitment of the kind involved in avowals and promises is so common that in many contexts we have to go to some trouble to indicate that we are just reporting on ourselves rather than making an avowal, or just making an avowal rather than making a promise. Thus to say "I believe that *p*" or "I intend to do *X*" will usually be taken as an avowal rather than a report, and to say "I'll be there" will often be taken as a promise rather than an avowal of intention. In order to signal that you are just making a report in the first case, just making an avowal in the second, you will have to resort to special phrasing. In the first case, you will have to say something like "I think that what I am inclined to believe or prefer or choose is such and such but I may, of course, be mistaken about my attitudes." And in the second, you will have to say: "My intention is to be there but I could yet change my mind."

It is because we understand the meanings of words in common, can look to the demands of reasons in determining what to affirm verbally, and can use words to speak for ourselves in a commissive mode that we count, in my terminology, as conversable beings. We interpret ourselves in words, using

them in a conventional sense, and we give that interpretation a unique first-personal authority in our dealings with others by investing the utterances with commissive force. We put our reputation on the line, using the words to invite others to see us in this and that manner: to see us as agents who believe such and such, desire so and so, and intend to do this or that.

We are conversable insofar as our use of words in this interpretive, commissive manner means that we are subjects with whom others can do business. We let them see that we form our attitudes on the basis of the reasons that are relevant, by common consent, to those attitudes. And we give others a powerful basis—one rooted in the interest we have in forming and maintaining certain relations with them—to expect that as we assert things to be, so we will prove to believe them to be; that as we assert that we shall make things be, so we will prove to intend them to be; and so on. Or at least we do this when we do not seek to deceive or manipulate others, trying to get away with masking our real beliefs and intentions. I put aside that complication for the sake of simplicity, although it is clearly relevant in the personal and, to anticipate, in the corporate case.

Corporations Are Conversable Agents

Having spelled out what conversability involves, it should be clear that the agents constituted by corporations, or by any corporate bodies, are conversable in much the manner of individual human beings. The spokespersons who speak in the name of the corporation do not speak just as reporters, as if their job was to take a census on the attitudes of members about any issue they address and then report on what those attitudes seem to be. They speak with the same sort of authority that any one of us assumes when we speak as individuals for ourselves. Where does this authority come from? Very simply, from the fact, long registered in law, that the other members of the corporation ascribe that authority to them, implicitly or explicitly committing themselves as individuals to rally behind the words of their spokespersons on any relevant issue; they treat those words as expressions of attitude that they have to live up to, on pain of corporate failure, in their actions as corporate members.

Thus suppose that the appropriate officers of a corporation enter a contract under which they commit the body to repay a debt on a certain schedule and at a certain rate of interest. Convention and law rule out the possibility that certain members of the corporation—say, a disaffected number of shareholders—might claim that the officers got their intention and promise wrong and that the company is not obligated to meet the terms of the debt. Insofar as the officers acted under the rules of the corporation in accepting and entering the contract—the rules implicitly or explicitly endorsed by members—they made

it the case that the corporation formed the required intention and promise; if you like, they made up the mind of the corporation.

You and I have a powerful motive to authorize the words we use in avowal and promise, since this enables us to gain credibility with others and form productive relationships with them. And in the same way the members of a corporate body such as a corporation have a powerful motive to authorize the words of their spokespersons when they make avowals and promises in the corporate name. Unless they authorize those words, being prepared to live up to whatever the words require of them—being prepared, in our example, to go along with the agreed repayment of the debt—they cannot expect to gain the benefits of incorporation. They cannot expect, for example, to be able to persuade any other individuals, or indeed any other organizations, to rely on their corporate word and to cooperate with them in any projects or arrangements that they propose.

That corporations and other such bodies are conversable agents means, in quasi-legal parlance, that they are persons. Not persons in a moral sense—not persons in a sense in which this involves a moral status—but persons in the functional sense of agents who can speak for themselves in a commissive manner. When the idea of the corporate body came to fruition in the high Middle Ages, it was expressed in the claim that such bodies are examples of what Pope Innocent IV cast in a document issued in 1246 as a *persona ficta*, a fabricated person (Kantorowicz 1997). Understanding personhood in a moral sense, the theologians generally took this to mean that the corporation is only a pretend person: a fabrication (Eschmann 1946). But understanding personhood in a functional sense, legal and political thinkers argued, ultimately with more influence, that it meant that the corporation is an artificial person, fabricated in the distinct sense of being made by human hands.

On this interpretation, medieval and later commentators cast the corporate body as a real person but, unlike a natural person, not the bearer of an immortal soul (Woolf 1913; Canning 1980). Thus, as Innocent IV insisted, the corporate body—for example, a body like the University of Paris, the addressee of his 1246 document—could not be excommunicated and consigned for eternity to hell's fires. But, so the lawyers argued, it could be held responsible in law and subjected reasonably to more mundane penalties. I return to the theme in the next section, when I discuss the connection between the conversability of agents and their fitness to be held responsible for what they do.

Corporations Have Minds of Their Own

Before moving on to that section, however, it will be useful to add one further element to our picture of corporate bodies as conversable agents and functional persons. While Thomas Hobbes (1994, ch. 16) also regarded

corporate bodies as conversable agents—persons, as he put it, that speak for themselves—he nevertheless held that in comparison with natural persons, they come into existence only by fiction (Skinner 2010). What I take him to mean is not just that they are artificial rather than natural persons—that would have gone without saying—but that they are persons or agents without minds of their own: agents that borrow the attitudes for which they stand from elsewhere rather than forming them for themselves.

This claim is clearly true of the corporate body with a single, dictatorial spokesperson—in effect, a Hobbesian monarch (Pettit 2008; 2014)—since that body borrows from its leader the voice behind which members rally and acts in expression of the leader’s mind, not a mind of its own. But Hobbes thinks that all corporate bodies, even ones with a non-dictatorial constitution, are dependent in the same way on recruiting from elsewhere the voice it enacts and the mind it expresses.

Hobbes argues that any non-dictatorial corporate body must authorize the voice of a committee, whether an elite committee or a committee of the whole, and that this voice is determined independently in much the same way as the voice of a dictator. He holds that the committee voice has to be fixed by majority voting, qualified to allow for ties. And, treating this as a voice that is algorithmically determined by majoritarian procedure, he thinks of the group that follows it as an entity that does not have a mind of its own. In following majority voting the corporate body expresses a mind that it recruits, not now from an individual spokesperson, but from a more or less mechanical procedure. And so it does not make up its own mind, as an individual might make up his or her mind, but helps itself to a mind that is independently fixed.

The final element to add to the picture is the observation that Hobbes is mistaken in this line of thought. If a corporate body is to be conversable, and to succeed in getting others to do business with it, then a minimal constraint on the voice it follows, and the mind it expresses, is that it should be sensitive to requirements of consistency; it should display consistency or at least be responsive to evidence of inconsistency, say evidence manifested in the complaints of others. But it turns out that the majority voice of a group on a set of interconnected judgments—even a set as simple, say, as “p”, “q” and “p&q”—is liable to be un-remediably inconsistent, as the discursive dilemma shows (Pettit 2001a, b).⁵ And it transpires, more generally, that any equally mechanical way of constructing a group voice is subject to similar problems, as the literature on judgment aggregation has established (List and Pettit 2002; List and Polak 2010; List and Pettit 2011).

⁵ The discursive dilemma is a generalized version of the doctrinal paradox in legal theory (see Kornhauser and Sager 1993; 2004).

The discursive dilemma is readily illustrated for our simple example with “p,” “q,” and “p&q.” Suppose that a group of just three individuals A, B, and C, each with a consistent set of views about those propositions, relies on majority voting to form a group view. It is perfectly possible for the majority voice to support an inconsistent set of views, with the majority supporting each of the simple propositions, for example, but rejecting the conjunction. Table 1.1 illustrates that possibility.

In order for a corporate body to be conversable the members cannot rely on any mechanical rule to determine the body’s views; they must institute a system for gaining feedback on the results generated by the protocols they follow, and they must find a way of correcting for any inconsistency produced. The simplest variation on majority voting that would achieve this is the straw-vote procedure (List and Pettit 2011, ch. 2). This would require the members to check after every vote on whether the view endorsed in the vote is consistent with what has already been endorsed and to look for a compromise in the event that it hasn’t. Following such a procedure, A, B, and C might agree that as a group they should endorse “p&q,” rejecting the majority judgment on that particular proposition. Or of course they might form a different compromise and reject the majority judgment on one of the other propositions instead.

By making the move toward any such compromise the members would cease to look elsewhere to find a voice to follow and a mind to express. They would construct the voice of the corporate body in a way that makes it relatively autonomous from the individual voices of members. And they would make up the corporate mind that their actions as members then go on to express. This holds, not just for the straw-vote procedure of incorporation, but for the myriad alternatives under which a group could allow feedback on the consistency of the attitudes its protocols support, seek to correct for failures, and establish itself as a conversable entity.

Just to illustrate the alternatives possible, those procedures might operate on the outputs, not of voting by members of the group as a whole, majoritarian or otherwise, but on the outputs of different committees on different issues, or indeed on outputs that require the support of distinct committees. And of course corrective adjustment itself might be conducted in different ways: by the membership as a whole, or by a particular committee, or by different

Table 1.1. A discursive dilemma

	Question 1: p?	Question 2: q?	Question 3: p&q?
Person A	P	Not q	Not p&q
Person B	P	Q	p&q
Person C	Not p	Q	Not p&q
The majority	P	Q	Not p&q

committees in different areas. The possibilities are legion and include the sort of mixed constitution—or “mixarchy,” as Hobbes called it—that Hobbes himself was anxious, for anti-republican reasons, to resist in the case of the political state (Pettit 2008). As exemplified in the United States, for example, this would require laws to be supported by two houses of Congress as well as the President and would allow the Supreme Court to strike any law down on grounds of being inconsistent with constitutional and other commitments.

PREMISE 2: CONVERSABLE AGENTS ARE FIT TO BE HELD RESPONSIBLE

Holding Agents Responsible

What does it mean to hold agents responsible for doing something? It certainly does not just mean claiming that they were causally at the origin of the effect in question. If it meant this, then we could be said to hold a hurricane responsible when we identify the damage that it did to our neighborhood. There may be a causal sense in which we hold things responsible for their effects, but it is not the sense that is relevant to the question about whether we should hold certain agents responsible: whether they are fit to be held responsible for the action in question.

Perhaps what we mean by holding an agent responsible is illustrated by the regulative stance we take, not to an uncontrollable force like a hurricane, but to our pet dog when we blame it for the wet patch on the carpet or praise it for seeking to go outside when it needs to relieve itself. In this case, by contrast with the case of the hurricane, we recognize that we can train the dog to behave in a certain way and that the best means of doing this is to sanction it by rewards and penalties for how it behaves, praising and petting it when things go well, or wagging our finger threateningly at it when things go badly. Perhaps sanctioning the dog for behaving or not behaving to plan can be usefully taken to exemplify what it means to hold an agent responsible.

This is not satisfactory either. Were I to try to influence you in this conditioning way, the success of the exercise might well depend on my hiding the nature of what I was doing from you; making the project clear would be liable to elicit resentment and reactance in you.⁶ That is why we laugh at the cartoon in which the boss leans across his desk and says to a shrinking employee: “They tell me, Jones, that you respond well to intimidation.” But

⁶ Finding it objectionable, you would be likely to prove defiant and counter-suggestible in response to a transparent attempt to condition you, thereby undermining the regulative point of the exercise (Brehm and Brehm 1981; Grabosky 1995).

when I hold you responsible, what I am doing is capable of being made manifest between us and will normally be transparent in that way. And so what I am seeking cannot amount just to the conditioning form of influence that I look for with the dog.

The reason why the sanction-based picture of holding responsible fails the transparency test, plausibly, is that it does not treat the subject who is held responsible as a conversable person. You will have the capacity as a presumptively conversable subject to form attitudes that correspond to the words you are led to endorse under the pressure of relevant reasons, whether of evidence or of interest. And if I want you to behave in a certain manner, the saliently appropriate way of achieving this effect is to appeal to reasons to support the line that I want you to take; only that approach takes account of the very special sort of agent that you are. To seek a conditioning effect, akin to the effect I seek with the dog, would be to try to get in under the radar of conversability. It would be to try to influence you, not as a person in the functional sense of the term, but rather as a slave of stimulus and effect.

This observation points us toward an attractive account of what it is to hold you or any other agent responsible, at least in the special situation where you have spoken about your beliefs, intentions, and the like prior to action. It is to assume that in issuing such words, you were avowing attitudes and promising actions. And then it is to hold you to the expectation, elicited by the utterance of those words, that you will act accordingly: you will live up to those words, proving to display the required attitudes and to perform the required actions.

If you meet that expectation, then I will confirm you in the ascription of conversability and, other things being equal, reward you with my approval, reliance, and cooperation. If you fail to meet that expectation, however, I will respond in either of two ways, one of which puts your capacity to cooperate in question, the other your reliability in exercising the capacity. Thus I may put your very conversability in question, wondering if you have the capacity to understand and honor the import of the words you utter. Or, to take the more likely possibility, I may take you to have that capacity and to have failed to exercise it—to have failed to match words and deeds—in which case, as we generally put it, I will rebuke or censure you. I will rebuke or censure you, at any rate, in the absence of credible and acceptable excuses for the failure. These are excuses that avowals and promises would still leave in place and include ignorance of what was at stake that you were not in a position to notice; pressures, as in prospects or threats of retaliation, that would have made it heroic to live up to the words; or obstacles like a broken leg that got in the way of your being faithful to the words (McGeer and Pettit 2015).

This explanation of what it means to hold you responsible is confined to the sort of case where you utter words in token of what you claim to think or be ready to do and assume responsibility for making your deeds match those words. But I may also hold you responsible in the more general case where you

do not explicitly assume responsibility for it. This is the sort of situation where, as a matter of common belief between us, you recognize the expectation on my part that you will conform to common standards, do nothing to cancel that expectation, and then fail—or, as it may be, manage—to meet it. That this is a matter of common belief means that we each believe it, we each believe that we each believe it, and so on in the usual hierarchy (Lewis 1969).

In this more general case, as in the other, you communicate your mind in a commissive way to me and others, but you do so by virtue of what you do not say rather than by virtue of what you do say. Following this pattern, we manifestly or overtly hold you to expectations of non-violence, non-deception, non-coercion, and the like, even when you do not explicitly commit to these. And if you fail to meet those manifest expectations we will hold you just as responsible as if you had explicitly made suitable avowals and promises. You may incur responsibility, even when you do not explicitly assume it.

Fitness To Be Held Responsible

There are three conditions that must be satisfied, intuitively, if you are fit to be held responsible for doing or not doing something, X.⁷ They correspond, broadly, with the conditions of grave matter, full knowledge of the guilt, and full consent of the will that were set out in the traditional catechism.

- *Grave matter.* You must have had a presumably non-trivial choice between doing X and not doing X or between doing X and more specific alternatives.
- *Full knowledge of the guilt.* You must have been in a position to gauge standard, presumptively common expectations and to make a normative judgment about the merits of the options, as registered in those expectations.
- *Full consent of the will.* Whatever direction that judgment assumed, you must have had the capacity to choose as that judgment would have required you to choose; and to do this, moreover, because this is what the judgment required.

On this account of fitness to be held responsible for a given deed, it should be clear that you are going to be fit to be held responsible just insofar as you are

⁷ Notice that as Harry Frankfurt (1969) argues, you may still be held responsible for doing X in a case where you thought you had the alternative of not doing X but were deceived in the matter; unbeknownst to you, you would have been forced to do X, even if you had chosen not to do it. But even in this sort of case you did have two alternatives open. You could have done X voluntarily or you could have tried not to do X. You will choose that second option if you take steps that trigger the force which leads you to do X involuntarily.

a conversable agent. You must have the generic, unimpaired capacity associated with conversability to live up to the expectations you explicitly or implicitly endorse. And that capacity must be unimpeded in the specific choice on hand; it must not be temporarily blocked by ignorance or pressure or obstruction of some kind. Let the generic capacity be absent and you will be exempt from being held responsible from forming or living up to any commitment, explicit or implicit. Let the specific capacity be absent and you will be excused by the factor that impedes it, assuming that you cannot be held responsible on an independent basis for letting that factor get in the way (Gardner 2007). But let both the generic and specific capacity be present and you are firmly on the hook in the case of failure, safely on the podium in the case of success. You will be deserving, as we say, of blame in the one case, praise in the other.

The third condition in this account may seem to require a non-naturalistic free will: a contra-causal capacity, whatever the other factors at work in motivating your decision for or against X, to have set those aside and injected your act of will into the causal chain. Suppose that your capacity to act as your normative judgment requires involves nothing like that sort of free will but consists in a disposition, elicited under natural and cultural pressures, to respond reliably to the normative considerations present; suppose it means that if you do not respond appropriately then that is just a fluke (Smith 2003). If your capacity amounts to nothing more than this, can it really support our holding you responsible, treating you as deserving of censure or praise? The following challenge suggests not.

Suppose you have the generic and specific capacity to do something good, X: you are generally conversable, having the unimpaired capacity to act on normative judgments, and there is no excusing factor that impedes the exercise of that capacity in choosing between X and the alternative. And suppose, as is surely possible, that you fail to exercise your capacity and deliver X, whether because of failing to make the right judgment or failing to act upon it. If your capacity consists just in a disposition to respond reliably to the normative considerations supporting X, then presumably the failure to display that capacity must have been due, at least in the last analysis, to some chance perturbation: perhaps a neural misfiring in your brain. But if the failure was the result of such a misfiring, how can I hold you responsible for it? How can I say, as we do say in such a case, that you could have done otherwise; you could and indeed ought to have responded appropriately to the reasons in favor of doing X? Shouldn't I rather acknowledge that your failure was due to something going awry in the neural works, in which case it would seem inappropriate to lay any blame at your door?⁸

⁸ For a further development of this challenge, and a more elaborate response, see McGeer and Pettit (2015).

No, I should not dismiss your failure in this way. However transparent and acknowledged, holding you responsible is a regulative practice in which I enjoin you, by virtue of manifesting a certain expectation, to perform in a corresponding manner. And a little reflection on the point of any regulative practice explains why in the case envisaged I should insist that you could have done otherwise and that you are subject therefore to censure.

The point of a regulative practice lies in the presumptive fact that enjoining you to perform appropriately is likely to increase the chance of your doing so. Or at least it is likely to do this insofar as it implies a penalty for non-performance: it takes the form “Do this, or else . . .” Any injunctive practice will presuppose a generic capacity to comply, as in the presupposition that you are a conversable subject, if it is to serve a regulative function. And any such practice will recognize certain factors—certain excusing factors, as they will seem—such that there is no point in penalizing a failure when it is produced by such a factor. The assumption is presumably that there is little or no hope, in the presence of such excusing factors, of increasing the chance of the person acting as enjoined: the factors in question put the agent beyond the regulative reach of the practice and count for that reason as excuses.

If this is correct, then it should be clear why we should stick to our penalizing instincts in the sort of case envisaged, where you fail to live up to our shared expectations, despite having a specific capacity to have done so. I may freely admit that your failure must have been due ultimately to some misfiring in your makeup—say, some neural glitch—but to give up on censuring you in any case where this is plausible would be to deprive the practice of holding you responsible of any regulative point. It is important with a regulative practice to identify the general capacity it presupposes, and to identify those excusing factors that let the agent off the hook: those factors that make the agent incapable of being regulated profitably. But it would make no sense whatsoever to let the agent off the hook whenever there is some presumptively naturalistic but non-excusing explanation for the failure. It would be to deprive the practice of the regulative role that, by hypothesis, it is meant to serve.⁹

Conversability, Responsibility, and Corporations

What these considerations show is that to be an unimpaired, conversable agent like you or me is to be fit to be held responsible for various actions, in

⁹ The position sketched here, following McGeer and Pettit (2015), steers a middle course between those on the one side who hold that being fit to be held responsible presupposes that you have a contra-causal free will and those on the other—now an increasing number—who argue that it only presupposes a general capacity to be responsive to the judgments of others: a capacity that implies that the attitudes you have speak to the sort of person you are but not in any demanding sense that you have done otherwise than you did.

particular those in the exercise of which there is no excusing factor that impedes our conversable capacity. But the fact that this is true at the level of conversable subjects like you and me suggests that it should also be true at the level of conversable corporate bodies, including corporations.

Like individual human beings, corporations routinely face decisions on non-trivial matters; in particular, on matters that involve the welfare of ordinary people. Like individual human beings, they are capable as conversable subjects of forming normative judgments on the relative merits of the options before them in any instance. And like individual human beings, they are capable as conversable subjects of letting the normative judgments they endorse constrain what they actually do. Thus we are all in a position to call such a body to account for what it does. More than that, indeed, most of us are in the habit of doing this in a more or less informal way. We would all tend to blame a company for negligence if it put toxic food on the market, for example, or for wrongdoing if it hid fatal design flaws in the motor cars it produces.

But could a corporation not respond to our calling it to account in these ways by maintaining that it was unable to form suitable judgments, or unable to act on the judgments it forms? Could it not claim in that respect to be a special, restricted sort of conversable agent, not an agent like you and me? No, it could not.

Suppose, to take one possibility, that a corporation announced that it was unable to make normative judgments on an issue like whether or not it is wrong to put toxic food on the market or to hide potentially fatal design flaws in the cars it manufactures. In order to make this claim, the corporation would have to have the concepts of what is right and wrong, and the associated capacity to make corresponding normative judgments. And so in making the claim it would presuppose that the content—the proposition asserted—is actually false. It would be guilty of what is known as a performative contradiction. The claim would presuppose that something is so—that it, the corporation making the claim, has a certain capacity—but assert in contradiction of this presupposition, that actually it, the corporation, does not have that capacity.¹⁰

Or suppose, to take a less radical possibility, that the corporation announced that while it was capable of making such normative judgments, it did not have the capacity or wherewithal to ensure that it lived up to them. It did not have the resources to monitor and police the decisions of its employees so as to reduce the probability of putting toxic food or faulty cars on the market. This claim might not be inconsistent with any presupposition it strictly makes, but it is inconsistent with the manifest fact that even if a corporation lacks such monitoring and policing resources, in most relevant

¹⁰ It is akin to the claim: “P; but I don’t believe it.” The making of this purportedly sincere claim presupposes the belief that p, as our example presupposes certain conversable capacities, but the content of the claim is that this presupposition is unfulfilled.

areas it will have institutional resources sufficient to establish the sort of disciplinary regime required. If the corporation has the capacity to give itself the capacity to live up to the judgments that it is wrong to put toxic food or faulty cars on the market, then it has the capacity to live up to such judgments, period. It might be able to invoke the need for more time to excuse a failure to live up to the judgments on the first or second occasion when it endorses them. But it will have no excuse for not living up to them over the longer haul.

CONCLUSION

My first premise was that corporations are conversable agents—persons, in a functional sense of the term—and my second was that conversable agents are fit to be held responsible. This supports the conclusion that corporations, and indeed corporate bodies more generally, are fit to be held responsible. In particular, since this is where the sharper implications arise, they are fit to be held responsible for actions that breach the common, shared expectations to which we are disposed to hold such entities.

Three points are worth stressing, however, in qualification of this conclusion. The first is that to be prepared to hold corporations responsible is consistent with being prepared at the same time to hold individuals responsible for the things they do in the corporate name. We may think that a corporation should be held responsible for distributing contaminated food, for example, and at the same time argue that certain employees who were negligent in the exercise of their corporate roles should also be held responsible for the part they played. The corporation allows or programs for any such failure, insofar as its procedures make it possible or probable. The individuals who are negligent in the exercise of their roles are in a position where they could have refused to play by the corporation's rules but nonetheless did so. Both are fit to be held responsible in their particular domains of control.

The second point to note is that even if individual members may be held responsible for the part they play in corporate malfeasance, there is still an important point to holding corporations responsible too. This appears in the fact that often the individuals who play a damaging part in a reprehensible corporate act have reasonable excuses for their failure. They may claim that they were not properly informed about the effects of their actions, that they were conforming to widely established precedents, that they operated under constraining expectations, or that they lived in fear of losing their jobs. Suppose we allow individual malefactors off the hook on such grounds, as we often should do. That means that unless we are prepared to blame the corporations they serve—and in effect the memberships who sustain those bodies—as their fitness to be held responsible allows us to do, there will be an

avoidable shortfall in the regulatory effects that our responsibility practices are generally designed to achieve. We will leave a loophole for people to incorporate for socially harmful but selfishly rewarding ends, and to do so with relative impunity.

The final point to note is that I have been concerned in this essay mainly with the issue as to whether it is appropriate to hold corporations responsible within our ordinary practices of normative assessment and censure. I believe that their fitness to be held responsible in this way argues for the appropriateness of holding corporations responsible in the criminal law—and not just, for example, in the law of torts—but I have not addressed the issue of criminal responsibility in this particular paper. It is a topic that I would prefer to leave to those better positioned to pronounce on matters of legal and institutional design (Fisse and Braithwaite 1993; Laufer 1994).¹¹

REFERENCES

- Anscombe, G.E.M. (1957). *Intention*.
- Brehm, Sharon S. and Jack W. Brehm (1981). *Psychological Reactance: A Theory of Freedom and Control*.
- Canning, J.P. (1980). "The Corporation in the Political Thought of the Italian Jurists of the Thirteenth and Fourteenth Century," *History of Political Thought* 1: 9.
- Ciepley, David (2013). "Beyond Public and Private: Toward a Political Theory of the Corporation," *American Political Science Review* 107: 139.
- Clark, Austen (1994). "Beliefs and Desires, Incorporated." *Journal of Philosophy* 91: 404.
- Dennett, Daniel C. (1987). *The Intentional Stance*.
- Eschmann, I. Thomas (1946). "Studies on the Notion of Society in St Thomas Aquinas, 1. St Thomas and the Decretal of Innocent IV Romana Ecclesia: Ceterum." *Medieval Studies* 8: 1.
- Evans, Gareth (1982). *The Varieties of Reference*.
- Fisse, Brent and John Braithwaite (1993). *Corporations, Crime and Accountability*.
- Frankfurt, Harry G. (1969). "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66: 829.
- French, Peter A. (1979). "The Corporation as a Moral Person." *American Philosophical Quarterly* 16: 207.
- Gardner, John (2007). *Offences and Defences: Selected Essays in the Philosophy of Criminal Law*.
- Grabosky, P.N. (1995). "Counterproductive Regulation." *The International Journal of the Sociology of Law* 23: 347.

¹¹ I am grateful for a range of helpful comments received when this paper was presented at a conference at INSEAD in December 2013 and to the legal theory workshop in Columbia University, New York, September 2015.

- Grantham, Ross B. (1998). "The Doctrinal Basis of the Rights of Company Shareholders," *Cambridge Law Journal* 57: 554.
- Hess, Kendy M. (2014). "The Free Will of Corporations (and Other Collectives)." *Philosophical Studies* 168: 241.
- Hobbes, Thomas (1994). *Leviathan*, ed. Edwin Curley.
- Kantorowicz, Ernst H. (1997). *The King's Two Bodies: A Study in Mediaeval Political Theology*.
- Kornhauser, Lewis A. and Lawrence G. Sager (1993). "The One and the Many: Adjudication in Collegial Courts." *California Law Review* 81: 1.
- Kornhauser, Lewis A. and Lawrence G. Sager (2004). "The Many as One: Integrity and Group Choice in Paradoxical Cases." *Philosophy and Public Affairs* 32: 249.
- Kraakman, Reiner, et al. eds. (2009). *The Anatomy of Corporate Law: A Comparative and Functional Approach*, 2nd ed.
- Laufer, William S. (1994). "Corporate Bodies and Guilty Minds." *Emory Law Journal* 43: 647.
- Lewis, David (1969). *Convention*.
- List, Christian and Philip Pettit (2002). "Aggregating Sets of Judgments: An Impossibility Result." *Economics and Philosophy* 18: 89.
- List, Christian and Philip Pettit (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*.
- List, Christian and Philip Pettit (2012). "Symposium on Group Agency: Replies to Gaus, Cariani, Sylvan, and Briggs," *Episteme* 9: 293.
- List, Christian and Ben Polak (2010). "Symposium on Judgment Aggregation," *Journal of Economic Theory* 145: 441.
- McGeer, Victoria and Philip Pettit (2015). "The Hard Problem of Responsibility," in *Oxford Studies in Agency and Responsibility*, vol. 3, ed. David Shoemaker.
- Orts, Eric W. (2015). *Business Persons: A Legal Theory of the Firm*. rev. paperback ed.
- Pettit, Philip (2001a). "Deliberative Democracy and the Discursive Dilemma," *Philosophical Issues* 11: 268.
- Pettit, Philip (2001b). *A Theory of Freedom: From the Psychology to the Politics of Agency*.
- Pettit, Philip (2003). "Groups with Minds of their Own." In *Socializing Metaphysics: The Nature of Social Reality*, ed. Frederick F. Schmitt.
- Pettit, Philip (2007). "Responsibility Incorporated." *Ethics* 117: 171.
- Pettit, Philip (2008). *Made with Words: Hobbes on Language, Mind and Politics*.
- Pettit, Philip (2009). "Corporate Responsibility Revisited." *Rechtsphilosophie & Rechtstheorie* 38: 159.
- Pettit, Philip (2014). "Group Agents are not Expressive, Pragmatic or Theoretical Fictions," *Erkenntnis* 79 supp. 9: 1641.
- Rønnegard, David (2013). "How Autonomy Alone Debunks Corporate Moral Agency," *Business and Professional Ethics Journal* 32: 77.
- Rovane, Carol (1997). *The Bounds of Agency: An Essay in Revisionary Metaphysics*.
- Skinner, Quentin (2009). "A Genealogy of the Modern State," *Proceedings of the British Academy* 162: 325.

- Smith, Michael (2008). "Rational Capacities, or: How to Distinguish Recklessness, Weakness and Compulsion." In *Weakness of Will and Practical Irrationality*, eds. Sarah Stroud and Christine Tappolet.
- Tollefsen, Deborah (2002). "Organizations as True Believers," *Journal of Social Philosophy* 33: 395.
- Velasquez, Manuel G. (1983). "Why Corporations Are Not Morally Responsible for Anything They Do," *Business and Professional Ethics Journal* 2: 1.
- Woolf, Cecil N. Sidney (1913). *Bartolus of Sassoferrato*.