dialectica

Vol 74 : 4
December 2020

The Mental States First Theory of Promising

Alida Liberman

doi:10.48106/dial.v74.i4.03

Online First

Alida Liberman. forth. "The Mental States First Theory of Promising." *Dialectica*. doi:10.48106/dial.v74.i4.03.

Please cite from the published version.
Check for updates here.

The Mental States First Theory of Promising

Alida Liberman

Most theories of promising are insufficiently broad, for they ground promissory obligation in some external or contingent feature of the promise. In this paper, I introduce a new kind of theory. The Mental States First (MSF) theory grounds promissory obligation in something internal and essential: the mental state expressed by promising, or the state that promisors purport to be in. My defense of MSF relies on three claims. First, promising to Φ expresses that you have resolved to Φ. Second, resolving to Φ commits you to Φing, all else being equal. Third, the norms on speech acts are determined by the norms on the mental states they express, such that publicly expressing that you are in a state subjects you to whatever commitments are normally incurred by being in that state, regardless of whether you really are in it. I suggest that this general approach might also explain how the norms on other sorts of speech acts work.

Philosophers have offered a variety of theories of promissory obligation, most of which ground promissory obligation in some external or contingent feature of the promise. In this paper, I sketch a new kind of theory, which instead grounds promissory obligation in something internal and essential to promise-making. This *Mental States First* theory of promising (or MSF) posits that the norms on promises are fixed or determined by the norms on the promises' underlying mental states; the mental states are "first" in the sense that they are explanatorily prior. My aim is to show that MSF should be taken seriously, because it is grounded in plausible assumptions, can accommodate a wider range of cases than can most theories, and fruitfully situates promising as part of a broader pattern of speech acts that behave in similar ways.

I motivate the search for a new theory of promissory obligation in Section 1 by pointing out that most theories face cases of apparently genuine promissory obligation that they cannot accommodate. In Section 2, I lay out how MSF

works. In Sections 3, 4, 5, I sketch arguments for several claims on which the success of MSF depends; if these claims hold, then MSF is a viable theory of promissory obligation. My argument in Section 5 highlights how the general MSF pattern holds for other speech act/mental state pairs, which suggests that a Mental States First approach might be productive for understanding the norms on other kinds of speech acts, as well.

1 Motivating MSF with Marginal Cases

Most theories of promissory obligation have no trouble explaining why we are obligated to keep our promises in paradigmatic cases, or those in which the promise is something the promisee wants and expects will occur, the promisor intends to bind herself to act as promised, harm would occur were the promise broken, and the promise occurs within some obvious social convention of promising. However, these theories struggle with explaining why we are obligated to keep promises in non-paradigmatic cases, which I call *marginal cases.*[1] These are cases in which there are good theoretical arguments—or a strong intuitive presumption with no good arguments to the contrary—that the case generates a genuine promissory obligation, but in which it is doubtful whether the theory can explain or accommodate the case as an instance of promissory obligation. Proponents of particular theories tend to wield such cases as dialectical weapons against each other (e.g. by arguing that view A is inadequate because it cannot capture case X, or that view B should be preferred because it can).[2] This approach presumes that theories should be able to explain every plausible case of promissory obligation, which makes sense as a desideratum for a good account; more theoretical breadth and explanatory power is usually better than less. Moreover, since promises

1 My use of "marginal" to describe these cases is not meant to imply that these cases only barely count as promises. Rather, I mean to convey that they are cases that ought to be considered genuine promises but that lack some of the core features of the most obvious and paradigmatic promises, which places them at the edges or borders of our practices of promising. Thanks to an anonymous referee for discussion of this point.

2 For example, Scanlon criticizes conventionalist views in this way, by arguing that they cannot accommodate a proposed case of promising without a social practice. And Scanlon's critics make similar sorts of arguments about his view, claiming that it fails because it is subject to a counterexample; see Cholbi (2002) and Southwood and Friedrich (2009). Of course, philosophers who argue in such ways typically also offer theoretical objections against the view they are targeting. But they generally assume that being unable to accommodate a case is highly problematic for a view.

appear to be a unified phenomenon, we want a unified account of them if possible.

It is important to note that a promise does *not* count as marginal for a theory simply because the theory implies that there are circumstances under which breaking the promise is morally permissible, all-things-considered. For it is widely accepted that promises yield pro tanto moral obligations, which can be overridden if excusing conditions arise (such as needing to break the promise in order to satisfy a more important conflicting obligation). When I claim that a case is marginal for a theory, I am claiming that the theory struggles to explain why the promise in the case has *any* moral force, or why there is even a pro tanto obligation to keep it.

In the rest of this section, I give a brief overview of how the most popular theories of promissory obligation are subject to marginal cases.[3] We can sort views with contemporary traction into three broad categories.[4] First are conventionalist accounts like that of David Hume—according to which we have reason to keep promises because it is bad for us (and our reputations as trustworthy people) if we do not—or John Rawls, who argues that failing to keep promises problematically free-rides on a valuable social practice.[5] Second are expectationalist accounts like that of T.M. Scanlon, according to which promise-breaking is impermissible because it involves violating expectations that you have raised in the promisee.[6] A third class of recently popular

3 This is not meant to be an exhaustive survey of theories of promissory obligation. I do not address views without much popular support or influence in the contemporary literature on promissory obligation, such as virtue-ethical approaches, intuitionist accounts, and Kantian deontology. Nor do I discuss act consequentialist views, according to which promises do not generate pro tanto obligations, and promise-keeping is morally required only if this leads to the best overall consequences. Of necessity, my discussions of the theories that I do address are sketchy and superficial; for more detail on how these (and other kinds of) theories of promissory obligation are subject to marginal cases, see Liberman (2015, chap. 2).

4 This follows Heuer's (2012) helpful classification.

5 See Hume (1739) and Rawls (1971). More recent conventionalist accounts vary greatly in their details; for example, Sheinman (2008) argues that promisors can give themselves social practice-based reasons to act by communicating the intention to give themselves such a reason. Hooker (2011) offers a rule consequentialist view that can be understood as a conventionalist view, which grounds promissory obligation in the rules that would lead to the best consequences were people to internalize them (see Liberman 2020 for further discussion of this view). A number of theorists have also proposed hybrid conventionalist/expectationalist views, which ground promissory obligation in expectations that can only be generated from within an existing social practice; see Kolodny and Wallace (2003).

6 Scanlon invokes a principle of fidelity that is justifiable on contractualist grounds; see (1990) and chapter 7 of (1998). Other views of the same sort ground promissory obligation in something

doi: 10.48106/dial.v74.i4.03

views are normative power accounts, which posit that we have the ability to change the normative situation directly and by declaration by exercising a normative power, through which we "change what someone is obliged to do by intentionally communicating the intention of hereby so doing" (Owens 2012, 4). Normative power theories claim that promissory obligation stems from an exercise of such power, and that these powers exist because they are valuable for us to have. In general, these views claim that X is an important feature of our normative lives, and that we can possess X only by having the ability to bind ourselves to each other through promising. In a transcendental step, they conclude that we therefore must possess the ability to bind ourselves to each other through promising, lest X be inaccessible.[7]

Each of these types of view—conventionalist, expectationalist, and normative power—gets something right, as each captures an important feature of paradigmatic promise-making, a feature which tells us something about why we are obligated to keep promises in many or even most cases. But each of these theories is also in some way incomplete; while they can explain why a moral obligation is present in paradigmatic cases of promising, they cannot explain the full breadth of cases in which we are obligated to keep our promises. For example, promise-breaking is often against one's self-interest because of the moral sanction it incurs, as Hume suggests. But we can easily imagine cases where this is not so; consider a traveler passing through a remote town who cons locals out of their cash, promises to pay them back, and disappears. Similarly, most societies have robust promising conventions that it would be unfair to free-ride upon, as Rawls claims. But we can imagine a successful

other than expectations; for example, Judith Jarvis Thomson (1990) argues that promise-keeping is required because the promisee is relying on the promisor, while Daniel Friedrich and Nicholas Southwood (2009) argue that promise-breaking violates the trust you have invited the promisee to place in you. Cholbi (2002) offers a contractualist view that grounds promissory obligation not in what the promisee *actually* expects, but in what she is *entitled to* expect; I argue in Liberman (2015, chap. 2) that this view cannot accommodate cases in which the promisee lacks moral standing to hold the promisor accountable.

7 Normative power theorists cash out what the normative power to promise consists in and what valuable feature of our normative lives it supports in a variety of ways. For example, David Owens proposes a normative power grounded in what he calls our "authority interest," or the interest we have in having a certain kind of practical authority over others. By making a promise, the promisor "give[s] the promisee the right to require performance of the promisor" (2012, 144); this serves our authority interest in allowing us to determine whether another person is obligated to act as she has promised, and in allowing others to make such determinations for us. Other normative power theorists speak instead of the ability to transfer rights (Shiffrin 2008) or create exclusionary reasons (Raz 1977).

promise in a state of nature without such conventions; Scanlon offers a case of strangers from different societies on opposite sides of a river, whose hunting weapons have fallen on each other's sides and who manage to successfully promise to exchange the weapons in the absence of a shared convention of promising. Conventionalist views cannot accommodate such cases.[8]

Similarly, Scanlon is right to point out that expectations matter; we are morally obligated not to mislead people and upset their expectations, or allow them to detrimentally rely on us and fail to follow through. But sometimes the promisee does not expect the promisor to perform; maybe the promisor is notoriously bad at keeping his word. As Berislav Marušić notes, expectationalist views entail that if the promisor does "not succeed in forming expectations in the promisee [...] she will thus fail to incur promissory obligations; her promise, if it is one at all, won't be binding" (2013, 305). But making a promise that doesn't generate expectations shouldn't get the promisor off the moral hook; it would be problematic indeed if you were justified in promise-breaking because the promisee assumed you were unreliable, and accordingly did not expect you to keep your promise.

Normative power accounts face their own marginal cases. For example, many people assume that there is a distinctively promissory obligation to keep promises we make to people who die before the promise can be kept. Promises to the dead are marginal for authority transfer views like Owens's. For there must be a person to whom the authority is transferred and in whom the authority continues to reside—but once the promisee is dead, no such authority bearer exists. Normative power theorists also presume that sincerely promising to Φ requires intending to obligate yourself to Φ. But this constraint leaves out some cases, as there seem to be cases of sincere promises that do not involve forming such an intention. As Thomas Pink (2009) argues, promisees are generally concerned not with whether the promisor is morally obligated to perform the promised action, but with whether the promisor *actually will* perform the promised action. And we frequently make promises without specifically intending to obligate ourselves in daily life.[9] I take it that you

8 More worryingly, Rawls's view fails to account for the directed nature of promissory obligation, as free-riding on a valuable social practice wrongs *all* of the participants in the practice equally, rather than wronging the promisee in particular. As Scanlon (1998) and Kolodny and Wallace (2003) both note, this is a major failure—although it is not a failure to accommodate a marginal case.

9 This is admittedly an overly quick dismissal of a popular view. I give a more detailed argument for this claim in Liberman (2015, chap. 4). See also Liberman (2015, chap. 5) for more about the sincerity conditions on promise-making.

doi: 10.48106/dial.v74.i4.03

can promise to Φ while merely *foreseeing* that your promise will incur an obligation but without specifically *intending* to incur such an obligation. For example, I can promise to take good care of a baseball glove I borrow from my brother with the intention of putting his mind at ease about getting the glove back in good condition, while merely *recognizing* (but not intending) that this promise will obligate me. I presume that you might also sincerely promise to Φ without even foreseeing that you will Φ (e.g. if you make a conditional promise and you believe that the relevant condition is extremely unlikely to occur), or to sincerely promise to Φ while being a nihilist about whether obligations can ever be incurred. Normative power views cannot easily accommodate such cases.[10]

The three kinds of theory of promissory obligation fail to explain why we are obligated to keep our promises in all—and only—cases of seemingly genuine promissory obligation. Conventionalist and expectationalist accounts risk being both over- and under-inclusive, because the sources of obligation that they point to are not *distinctively* promissory: that is, they stem not from the content or nature of the promise itself, but from some other, contingent feature of the situation (such as the context within which the promise occurs, or the downstream effects of the promise, which may not always be as expected). While promises often rely on social conventions, they need not do so, as we saw with Scanlon's weapon-exchange example. We could also have social conventions (such as an honor code) that led people to scrupulously keep their word for fear of social repercussion or free-riding without invoking promises at all. Likewise, not all genuinely binding promises raise expectations. And we sometimes raise others' expectations without making promises to them (e.g. by putting on your coat and raising my expectation that you are about to leave the building). Normative power accounts purport to be grounded in a feature of promises as-such (i.e. the intention to obligate oneself that is expressed by a promise), but the feature they identify is not essential. Although

10 Normative power theorists generally construe exercises of normative power as intentions to change the normative situation (e.g. through promising, consenting, giving, etc.). It might be possible to develop a normative power view according to which the interpersonal act of promise-making itself counts as a direct exercise of normative power, regardless of what intentions (e.g. to obligate yourself, or simply to act) accompany the act of promise-making. Such a view could get around the specific marginal case I am currently addressing. But it would require positing a brute normative power without an underlying explanation of how this normative power functions, which would not be as deeply explanatory or satisfying an account of how and why promises bind than MSF can offer. Thanks to an anonymous referee for discussion of this point.

forming an intention to obligate yourself and thereby transfer authority might be one way of creating a promissory obligation, it is not necessary.

This is not to say that theories of promissory obligation that are subject to marginal cases are without worth. To the contrary, maintaining valuable social practices and satisfying expectations matter morally. Considerations such as these often make it the case that a particular promise is morally obligatory to keep on independent grounds; people are often obligated to act as they have promised for reasons having to do with fairness, harm avoidance, and the like. But as already noted, such sources of obligation are neither maximally broad nor distinctively promissory. In the rest of this paper, I argue that there is another source of promissory obligation that is both broad enough to cover all cases, and distinctive of promising as such.

2 Introducing the Mental States First Theory of Promising

The existence of marginal cases motivates grounding a theory of promissory obligation not in the contingent and external circumstances of particular promises, but in something that all promises share, something that is internal and essential to the act of promising itself. One thing all promises have in common is that they involve a communicative act, usually a verbal speech act but sometimes a written or non-verbal act (for a nod or a stern look can successfully communicate a promise in the right context).[11] Whenever you perform such a communicative act, you *express* that you are in a certain mental state, by which I mean that you convey to your audience that you are in that state, and really are in it if your performance is sincere. If we can ground promissory obligation in the expression of a mental state that all promises share, we will have a maximally broad view.

This brings us to the following theory:

> Mental States First Theory of Promising (MSF). The obligation to keep promises is derived from the norms on resolutions, which are the mental state expressed by promising.

11 Could you have promises in which no direct communication occurs? In such cases of implicit promises, there is either successful indirect communication (e.g. if Sam and Chris agree to "go steady" in a context in which this conveys a commitment to monogamy, Chris breaks a promise by dating someone else), or there is no successful communication and no promise is made (e.g. if Sam and Chris start casually dating without any mutual presumption of sexual exclusivity, Sam does not break any implicit promise by dating someone else).

doi: 10.48106/dial.v74.i4.03

MSF claims that one who promises to Φ expresses that she resolves to Φ, and *in virtue of publicly expressing that resolution* is subsequently obligated to Φ, all else being equal.[12] I am proposing that the norms on promising can be explained by appeal to the norms on the mental state expressed by promising. This is a claim about explanatory priority; the norm on the mental state explains the existence of the norm on the speech act, because it fixes or determines what that norm is.

My argument for MSF is as follows:

1. Promising to Φ expresses that you have formed a resolution to Φ, conditional on the promisee's acceptance.

12 Downie (1985) articulates an account of promising that shares some structural similarities with MSF. Downie argues that "a promise is essentially a matter of pledging oneself" and that the promisee's "reliance and expectations are well-founded to the extent that they see that I already regard myself as obliged and they know me to be a man of my word [...]. To promise is always to state an intention in obligation-creating circumstances" (1985, 266). Downie argues that these circumstances are those in which "the intended projects have been made central and essential in one's total concerns. The self has been identified with the projects, and carrying them out has become not only a moral obligation of practical consistency but a strong moral obligation of honour or self-fidelity" (1985, 269). That is, Downie argues that there is a moral duty to maintain self-consistency about projects with which you have identified your will, and the promisor has "identified his will with the project" and "pinned his self on the future as described in his pledged actions" in a way that makes "keeping of the promise essential to the preservation of his personal integrity," such that "he will be diminished as a person if he breaks his word" (1985, 270). My view differs from Downie's in two primary ways. First, Downie's account accommodates only those promises that are deeply tied up with the promisor's sense of self, while MSF explains promissory obligation more broadly, including in cases that are not very important and that your sense of integrity is not wrapped up in. Second, Downie grounds the moral force of promissory obligation in the need to adhere to your important resolutions. While I agree that promissory resolutions generate moral obligations, I argue in Section 4.2 that this is because of the way in which they are conditional on the acceptance of the promisee, and not because there are general moral duties of self-consistency. Michael Robins (1984) offers an account that draws on action theory to ground promissory obligation in intention—specifically, in the intention that the assent of the promisee will obligate you to act as you've promised to. Robins begins with an "irreducibly normative" (1984, 12) notion of intention that binds the intending agent to act in certain ways in the future. He argues that vows are intentions about which the agent cannot change their mind, and that promises are the transferal of this "exclusionary mandate" about how one will act to the "normative control of another person" (1984, 120). Robins argues that this transfer of the exclusionary mandate to the promisee transforms the requirement to abide by one's vow into a moral obligation; various critics of Robins have argued that it is not clear what exactly this transfer consists in, how it occurs, or how it generates a moral obligation (see Cottingham 1985; Lemos 1987; Smith 1987). By contrast, MSF does not rely on any transfer (of an exclusionary mandate or any other right) to the promisee.

2. Resolving to Φ conditional on another person's acceptance rationally and morally obligates you to Φ (all else being equal).
3. In general, publicly expressing that you are in state X obligates you to act as X demands of you, regardless of whether you really are in state X.

---

4. Therefore, promising to Φ rationally and morally obligates you to Φ (all else being equal), regardless of whether you have really resolved to Φ.

The above argument depends on the following claims: (1) that promises express resolutions; (2) that resolutions rationally (and in some cases, morally) obligate you to act, all else being equal; and (3) the *determination claim,* according to which the norms governing the mental state expressed by a speech act (at least partially) determine what the norms on that speech act are, in both sincere and insincere cases. The latter premises are not derived from the former; these are all independent claims which together establish the conclusion. In the next three sections, I offer arguments to support each of these claims. Because I lack space to fully defend them at present, my argument is conditional: *if* these claims are true, then MSF is an appealing and viable theory of promissory obligation.

To clarify, MSF states that the mental state expressed by a promise—i.e. a resolution—is a necessary and ineliminable part of the explanation of why we are morally required to keep our promises. The mental state alone is not sufficient; an unarticulated mental state cannot generate interpersonal norms, and so the public expression of that state plays an essential role, as well. But the norms on the mental state are the original source of the obligation, and they determine what the norm on the publicly expressed mental state is: the norms on the promise derive from the public expression of the speech act of promising, and the norms on the public expression of the speech act of promising derive from (and are explanatorily downstream from) the norms

doi: 10.48106/dial.v74.i4.03

on the mental state conveyed by that speech act.[13] If the norms on this mental state were different, the norms on the speech act would be different, too.[14]

Because *every* promise expresses a resolution—that is, because every promise, regardless of whether the promisor is sincere or believed by the promisee, *conveys* that the promisor has resolved to act—MSF has broader applicability than do the theories discussed in the previous section, and can accommodate the marginal cases that these other theories cannot. When your highly unreliable friend tells you that he is going to pay you back the money you lend him, your expectations about repayment are not raised—but your friend does convey to you that he's resolved to pay you back (even if you doubt that he'll carry out this resolution). You can express resolutions in the absence of social conventions. A promise made to someone who later dies expresses a resolution in just the same way as a promise made to someone who remains living does. And one can convey that she has resolved to Φ without thereby communicating an intention to obligate herself. MSF is broader than the views discussed in the previous section because it derives promissory obligation from a core component of the speech act of promising itself, rather than from the content of particular promises or the various ways in which a promise interacts with the world (e.g. by creating expectations in the promisee, or being part of a social practice involving sanctioning).

13 The order of explanation goes from mental state to speech act rather than vice-versa, for there is a clear sense in which mental states are independent of and prior to the speech acts that express them, and in which speech acts are not independent of and prior to the mental states they express. One can properly and without any insincerity be in a given mental state without expressing it via a speech act. Since the mental states can properly function entirely independent of the speech acts, it would be strange to derive the norms on them from the norms on speech acts. But speech acts are not independent from mental states in the same way: they express that the agent is in a particular mental state, and it is problematically insincere to have the speech act without that mental state. This dependency makes it quite natural to derive the norms on the speech act from the norms on the mental state.

14 MSF states that you can obligate yourself at will (pending the promisee's acceptance) by publicly expressing that you have formed a conditional resolution to act. Does this make it a normative power view of the sort discussed in Section 1? It doesn't, because all theories of promissory obligation grant that promising involves deliberately creating a new obligation by doing something intentional (e.g. raising someone's expectations, or participating in a social practice, or conveying a certain intention). What makes normative power views distinct is the explanatory structure that they take (i.e. a transcendental argument grounded in the legitimate interest we have in being able to alter the normative situation by fiat). MSF proposes a different explanatory structure: the basis of the promissory obligation is underlying the norms on the mental state conveyed by the promise, and what the public expression of that state commits one to. Thanks to an anonymous referee for discussion of this point.

So far, I have illustrated what MSF can accomplish if it is true. My aim in the rest of this paper is to illustrate how plausible MSF in fact is. I begin by defending the claims that promises express resolutions (Claim 1). I defend claims (2) and (3) in subsequent sections.

3 Defending Claim (1): Promises Express Resolutions

What do you express or convey when you perform the communicative act that constitutes making a promise? That is, what mental state do you at least purport to be in when promising, and are you really in if your promise is sincere?[15] A natural idea is that promising to Φ expresses that you plan to Φ, and are serious about carrying out this plan. Someone who promises to Φ communicates that she really will Φ, even if she doesn't feel like doing so at the time, or a better option arises, etc. When I promise you that I will attend your show tomorrow, I'm telling you that I am going to be there, even if faced with barriers to action that might otherwise prevent me from going. If my promise is sincere, I really do have such a plan. If I am insincere, I express to you that I am serious about going to your show without actually being so committed.

How can we cash out this notion of a serious plan? We cannot appeal merely to desires, for desires aren't normatively committing in the way that promises are: promising to Φ pro tanto morally obligates you to Φ, but desiring to Φ clearly does not. Moreover, we often sincerely promise to do things that we don't desire to do; you can sincerely promise your department chair that you'll attend the next faculty meeting, even though this is not how you desire to spend your Friday afternoon.[16] Intentions are a mental state with more stability than desires, as Michael Bratman (1987, 18–20) and others have argued. If I intend to go to your band's show tonight, then I have settled the matter of what I am going to do. I should not continue to deliberate about it or revise my plans for no good reason. Similarly, promising you that I will attend the show settles the matter of what I am going to do. It would be inappropriate to continue deliberating about what to do or to revise my plans

15 I refer not to sincerity in the sense of being well-meaning or earnest, but to *communicative sincerity*. When A utters a speech act S that expresses state M, A is communicatively sincere if and only if A really is in state M.

16 You might have other desires that would be satisfied by going to the meeting (e.g. getting on the chair's good side). But we can imagine scenarios in which this is not the case.

doi: 10.48106/dial.v74.i4.03

unilaterally. Intentions might therefore be a decent candidate for the mental state expressed by promising.

However, mere intentions are not stable enough to capture the seriousness of the plan and the strong sense of commitment involved in promising. Suppose that I intend to go to your band's show tonight because I have nothing better to do. You attempt to solicit a promise from me to attend the show. I say, "I promise to be there. And as of now, I intend to go. But you should know that I'm not committed to refraining from revising that intention. My plans might change between now and then, especially if I get a better offer." Such a statement does not seem to be a genuine promise. This is because it is unproblematic for me to intend to go to your band's show tonight and then abandon my intention because a more appealing offer comes along. But such circumstances would not license my breaking a promise to you to see your band play.

This shows us that promises express something stronger and with more stability than typical intentions. Consider the special kind of intentions that we tend to form at the start of a new year—what we often call *resolutions*. These are particularly serious and stable intentions that we're strongly committed to, usually about important goals that we expect might be very difficult to attain. I take resolutions to be intentions that are especially robust or resistant to revision. [17] Resolutions are necessary when you plan to act and care about whether you do so but suspect that some temptation or other barrier to action (such as laziness, aversion to an unpleasant task, apathy, etc.) might cause you to abandon your plan were you not to bolster it somehow. Forming a resolution is one means by which you can bolster your plan and more effectively resist temptation.[18] It is plausible that when I promise to Φ, I express that I have resolved to Φ—that is, that I plan to Φ, and that I care enough about whether I do so that I will not reconsider or abandon that plan, even in the face of temptation, laziness, better offers, and the like. Unlike intentions, it is

17 The details of how we cash out resolutions do not matter, so long as there is some coherent notion of resolution that implies that we are irrational if we over-hastily revise or fail to act on our resolutions without a good excuse. In Liberman (2016), I argue that resolutions consist in an intention to act coupled with the desire not to reconsider that intention, and offer an objection to Richard Holton's (2009) closely related view according to which a resolution is an intention to act coupled with an intention not to reconsider.

18 I don't mean to claim that resolutions are the *only* effective means of resisting temptation. There are other means by which you can resist temptation, which can be more effective than resolution-making; the best way to refrain from the temptation to drink tonight might not be to form a resolution, but to lock the liquor cabinet and give the key to a reliable friend.

problematic to abandon a resolution because you no longer feel like acting, or because a better offer comes along. So a resolution-based account will not overgenerate cases of permissible promise-breaking in the way that an intention-based account would.

However, not every publicly expressed resolution counts as a promise; to simply announce in your presence that I have resolved to run a marathon is not to promise you that I will do so. This is because promises require a second party; the acceptance or uptake of the promisee is essential to making a promise. Promissory resolutions must therefore take account of promisee acceptance. They can do this if they are *conditional on* the acceptance of the promisee, in the way my resolution to go for a picnic tomorrow might be conditional on the weather being good. If I offer to promise to run a marathon with you, I convey that I resolve to run *on the condition that you accept* (and do not subsequently reject) my promissory offer. A valid promise is successfully created (and a pro tanto moral obligation generated) only if you accept my offer.

I propose that promises express a resolution to act, conditional on the acceptance of the promisee. However, we might worry that this account overgenerates cases of legitimate promise-making. For we can imagine publicly proclaimed resolutions that are explicitly conditional on someone else's acceptance or agreement, but that do not seem to count as promises. For example, suppose I tell my personal trainer that I resolve to lift weights with her five times a week, but only on the condition that she agree to work with me; she agrees. I have announced that I have resolved to train five days a week, conditional on her acceptance of this resolution, and she has accepted. But I don't seem to have made her a promise. How do we distinguish genuine promises from announcements that one has a conditional resolution?

The response to this worry is simple: it is plausible that a necessary precondition for validly promising is recognizing that you are making a promise in the first place. In general, successfully engaging in an intentional action that alters the normative situation requires a recognition of what you are doing, e.g. you must recognize that you are granting consent in order to successfully do so, and must be aware of the fact that you are transferring property in order to make a gift, etc. Similarly, it is plausible that I need to understand that I am making a promise in order to successfully do so. Likewise, it's plausible that the promisee must be aware of the fact that she is accepting a promise, as well. If the trainer does not *take herself to be accepting a promise* and thereby generating a valid promissory obligation when she says she'll work with me,

doi: 10.48106/dial.v74.i4.03

then what she is doing is agreeing with a resolution, and not accepting a promise.[19]

4 Defending Claim (2): Resolutions Commit You to Acting

4.1 *Resolutions and Rational Commitment*

Someone who resolves to Φ incurs a self-imposed, pro tanto, subjective rational obligation to Φ. I will refer to such obligations as *rational commitments.*[20] The easiest way to get a sense of what I mean by rational commitment is to consider the way in which holding one belief can commit you to holding another. Philosophers frequently talk about the ways in which our beliefs commit us, e.g. because Jack believes that only consequences are relevant for moral assessment, he is committed to believing that he ought to kill one person to save two. Such commitments can come apart from what you objectively ought to believe, all-things-considered: perhaps Jack should not believe that he ought to kill one to save two. A resolution commits you to acting in a

19 MSF proposes that promises are best understood as conditional resolutions, but we must appeal to the concept of a promise to distinguish which resolutions are promise-generating and which aren't. Is this problematically circular? It isn't, for MSF is not meant to be a descriptive account of what promise-making consists in or a tool for identifying which utterances count as promises and which don't. Rather, MSF is an account of the nature of promissory obligation and of the normative force of promises. MSF posits that this ultimately stems from the norms on the conditional resolution that is publicly expressed when you make a promise. Promisor and promisee must both take themselves to be participating in promise-making in order to determine which resolutions will play this role. But this awareness of what they are doing is not the fundamental normative mechanism and does not provide any deep explanation of how and why promises bind. And in general, it is unproblematic to appeal to the concept of X as one part of an explanation of the nature or normative force of X. For example, suppose I am offering an account of the nature and normative force of giving, according to which A gives X to B (in a moral, rather than a legal sense) if and only if A intends to transfer X to B in an irrevocable way. Some irrevocable transfers will fail to count as gifts—say, those that are made under duress and are perceived by A and B as threats. We can unproblematically state that A must intentionally conceive of themselves as giving X to B in order for X to count as a gift, and that gift-giving generally involves conceiving of oneself as making a gift vs. acceding to a threat. This doesn't diminish the explanatory force of the account of giving as irrevocable transfer; appealing to the *concept* of a gift is part of what cashing out the underlying *nature* and *normative force* of giving requires. Something similar is true for my account of promissory obligation: we must appeal to the concept of a promise to cash out the underlying nature and normative force of promising. Thanks to an anonymous referee for discussion of this point.

20 For more on the concept of commitments (as distinct from reasons and all-things-considered obligations), see Shpall (2013, 2014), as well as Liberman and Schroeder (2016).

similar way as believing that *p* commits you to believing the obvious and relevant consequences of *p*: someone who believes that *p* and that *p* entails *q* but does not believe that *q* when the question of whether *q* is salient fails to act on her rational commitments, and as a result her overall set of beliefs is not as complete and coherent as it should be. Similarly, someone who resolves to Φ at a particular time and then fails to intend to do so at that time because she readily abandons this resolution fails to act on her rational commitments, and as a result is not as effective a planning agent as she could be.[21] The rational obligation to act on one's resolutions stems from a broad demand for coherence in one's long-term plans; we are rationally committed to acting on our resolutions because this is essential for effectively carrying out our plans and acting in line with our important goals and values in the long term. To be clear, the obligation to fulfill your resolutions is not the same as a narrow requirement to be instrumentally rational or risk incoherence.[22] Rather, it is grounded in a broader demand for a more holistic sort of coherent planning agency. In order to meet our most important and difficult long-term goals, we must bolster ourselves against succumbing to temptation in ways that would undermine these goals. Resolution-keeping enables us to do this.

We can best illustrate how resolutions incur rational commitments—that is, how they impose subjective, pro tanto, rational obligations—with an example. Suppose you are generally hesitant to try new foods, and are deeply entrenched in the habit of eating pizza for lunch every day. You very much want to expand your culinary horizons, but are such a creature of habit that you are unlikely to do so unless you force yourself into it somehow. So you resolve to go to a Thai restaurant for lunch today, knowing that if you don't form this resolution you are likely to fall back into your pizza habit. When you leave your office to go eat lunch, you abandon your plan to go to the Thai place and head to the pizzeria instead, deciding that you might as well just eat pizza today, since it's easier for you to order from a familiar menu. Something is wrong with you in this picture; resolving to eat Thai food in order to expand your culinary horizons and then changing your mind without good reason is problematically

21 Time indexing is necessary to avoid over-generating cases: if I resolve to go to eat Thai food at noon, it's no problem that I haven't yet formed the intention to do so at 9 AM. But it would be problematic if lunchtime rolls around and I haven't formed such an intention.

22 For influential treatments of conditional normative requirements grounded in coherence as the basis of instrumental rationality, see Broome (1999), who articulates a wide-scope view, and Kolodny (2005), who articulates a narrow-scope view; see also Way (2010) for an overview of this debate.

doi: 10.48106/dial.v74.i4.03

self-undermining. It is irrational or incoherent to resolve to eat Thai food because you care about broadening your horizons, and then decide to stick with your pizza habit because it is easier. It's not that eating pizza every day is *independently* irrational; having pizza all the time is rationally permissible, if boring. Rather, it's that it is irrational to *resolve* to eat Thai food, and then abandon this resolution for no good reason.

Compare abandoning a resolution to eat Thai food to merely *desiring* to have Thai food and eating pizza instead. This is perfectly acceptable; failing to act on a particular desire is not irrational. Or compare it to the case in which you *intend* to have Thai food but don't really care about whether this plan changes. You do not display any irrationality if you change your mind because you suddenly have a craving for a sandwich; you've simply made a permissible change of intention on the basis of a change in desire. Changing your resolution on a similar basis is not so innocuous. When you abandon your resolution to go to the Thai restaurant for lunch, you are undermining your own endorsed goals and plans: you value culinary diversity, and adopt this as one of your goals, but do not succeed in attaining it.

We can best see how abandoning a resolution without good reason is irrational by comparing a pair of similar cases. Suppose that you and I each have a reason of strength X to eat Thai food and expand our respective culinary horizons. And suppose we each care about expanding our culinary horizons to the same extent. But only I take action about it: I *resolve* to eat Thai food, making it a part of my plan and adopting it as one of the concerns that I will focus on. If I change my mind and have pizza for lunch, I go wrong in a way that you do not when you have pizza. This is not to say that it is *never* permissible to abandon a resolution. Sometimes, there are weightier considerations in favor of revising a resolution than there are in favor of maintaining it—say, if you resolve to have Thai food and an old friend unexpectedly offers to meet you for lunch at the pizza place, or you realize that Thai food often contains ingredients to which you are allergic. However, resolutions rationally commit you to acting unless there is a good excuse in place.

4.2 *Resolutions and Moral Obligation*

It might seem prima facie puzzling to claim that the norms on promises can be explained by appeal to the norms on resolutions. I've argued in the previous section that resolving *rationally* commits you to acting. But promising *morally* obligates you to act, as well. Promissory obligations are also directed towards

another person in a way that most resolutions are not: a promise does not create a generic moral obligation to act in a certain way, but is owed *to the promisee,* who is uniquely wronged should the promise be broken. We can assess the directedness of an obligation by asking whether anyone is uniquely positioned to forgive (or has special standing to resent) the violator, where resentment is understood as a negative evaluative attitude that is appropriate only in response to a personal violation. For example, suppose a passenger on an airplane has a medical emergency. The flight crew asks if any medical professionals are on board. There is a doctor on board (of medicine, not of philosophy!) who could easily assist the passenger, but who fails to do so because she prefers not to miss any scene of her in-flight movie. The other passengers on the plane might criticize or blame the doctor for her callous behavior. But the distressed passenger has unique standing to resent or forgive the doctor in a distinctive way. The doctor owes it *to the distressed passenger* to help him; this is not a generic obligation to be beneficent, but an obligation owed to a particular person. How can a resolution yield a commitment that is both *moral* and *directed* in this sort of way?

I have three responses to this important question. First, I want to ease the explanatory burden on MSF. I am not trying to offer a complete or exhaustive account of all of the ways in which promises can morally obligate us, which means that MSF does not need to explain all of our intuitions about the robust moral force of promises. Rather, there are often *multiple* sources of moral obligation to keep any given promise, and these all contribute to the overall moral force of the promise.[23] These sources are the morally important considerations raised by the other theories of promissory obligation discussed in Section 1: I ought to keep my promise to you to pick you up from the airport because it will harm you if I don't, *and* because failing to do so would problematically free-ride on a valuable social practice, *and* because I have publicly conveyed to you that I have resolved to do so. This last source of obligation stems from MSF, and is present in all cases of promising.

Consider a pair of examples to illustrate how the moral force of a promise can vary (and will be stronger when other considerations are in play). First, suppose you promise your mother that you will not sell a treasured family heirloom that is in your possession. Assume that there aren't any strong reasons for you to sell it; you're doing fine financially, and the heirloom is

23 When there are multiple sources of obligation to keep a promise, the obligation is over-determined, and the moral force of it is stronger; see the case about reasons not to sell a family heirloom below.

doi: 10.48106/dial.v74.i4.03

not worth very much. The moral reasons you have to keep this promise are not exhausted by the fact that you've conveyed to your mother a resolution to hang on to the heirloom; they also stem from harm avoidance, family loyalty, the importance of maintaining family traditions, and the like. And so the moral force of the promise to refrain from selling the heirloom will be quite strong.

In cases in which no other reasons to keep a promise are present—that is, in which no expectation is formed, in which no trust can be violated, etc.—we see that promises generate comparatively weak moral obligations. Second, suppose that we are airplane seatmates stranded on the tarmac because of a flight delay. I ask you to promise me that you will write a negative review about the airline's customer service when you get off the flight, and you make the promise (and thereby resolve to do so, conditional on my acceptance). Assume that there are no independent sources of moral obligation to keep this promise—we are strangers, I will never know whether you write the review, doing so will in no way affect the airline's business, etc. You nevertheless seem to be pro tanto morally obligated to write the review. If you have an opportunity to easily write a negative review and you fail to do so without a good excuse, you will not be entirely as you ought to be. But neither will you be failing morally in a drastic way, for your obligation to keep the promise to me is not an especially important one. MSF purports only to explain this type of obligation: the relatively weak moral force that stems from the bare act of promise-making itself, including in marginal cases. This mitigates the explanatory demands on MSF, which needs to explain only this form of relatively weak moral obligation.

Second, the conditionality of the resolution on the interest of the promisee is what enables it to yield a directed obligation. Successful promise-making requires the acceptance of the promise by the promisee, and I've argued that promissory resolutions are conditional on such acceptance. Accepting a promise implies that you are interested in the promise being made, and (usually) that you are interested in it being kept. [24] It's plausible that resolutions can generally yield directed obligations when the interests of another

24 In typical cases, the promisee accepts the promise because she wants the promised action to occur. In deviant cases, the promisee might accept the promise for other reasons; Vera Peetz (1977) describes a case in which you accept your neighbor's promise to give you homemade jam not because you actually want her jam, but because you do not want to hurt her feelings by declining her offer. In this case, you are interested in the promise's being made—for this is necessary to spare your neighbor's feelings—even though you do not care whether it is kept.

person are directly involved in the core content of the resolution. Another person's interest in how you are going to act is normatively relevant, and in some circumstances—including promissory resolutions—this can ground a directed obligation.

We can illustrate how another person's interest in a resolution can yield a directed obligation by considering a series of cases. Suppose you resolve in secret to mow your elderly neighbor's lawn as a favor to him. In doing so, you incur a rational commitment. If you fail to keep the resolution without a good excuse for changing your mind, you will have acted irrationally, but will not have done anything morally problematic. If you inform your neighbor that you have resolved to mow his lawn, you will also incur a directed, expectation-based obligation to your neighbor to either mow the lawn or alter his expectations. And if you neither mow the lawn nor warn him in advance that you do not plan to do so, your neighbor is entitled to resent or forgive you for violating this expectation-based obligation. So too would be any third party who has a stake in the matter and whose expectations were similarly raised by your announcement of your resolution; if your neighbor's landscaper was present when you announced your resolution, she is likewise entitled to resent you for neither mowing the lawn nor warning her that you've changed your mind.

However, suppose you inform your neighbor that you have formed a resolution that is explicitly conditional on his interests—you say, "I've resolved to mow your lawn this weekend, so long as you want me to." In this case, it is somewhat plausible that you incur a directed obligation to your neighbor to mow the lawn. For your resolution conveys that you plan to mow his lawn, and that this plan depends not on *your* interests or desires, but on *his*. The conditionality of this resolution places the plan in his hands. If you don't mow the lawn and instead simply warn your neighbor ahead of time that you have changed your mind, you have perhaps done something morally problematic. It would not seem terribly out of place for your neighbor to resent or forgive you, in a way that would be odd if the resolution were not conditional on his interests. But if you warn the landscaper ahead of time that you will not be carrying out the conditional resolution as your neighbor wants you to, it *would* seem out of place for the landscaper to resent or forgive you; you have discharged your expectation-based obligation to her, and owe nothing more.[25]

25 There might be cases in which it is morally inappropriate for independent reasons to give the landscaper a warning instead of mowing the lawn yourself—say, if you know that the landscaper

doi: 10.48106/dial.v74.i4.03

It is even more plausible that a directed obligation is formed if your resolution is explicitly conditional not on your neighbor's *desire* that you mow his lawn, but on his active *acceptance* of your resolution to do so—that is, if you say, "I've resolved to mow your lawn, but will only do so if you remain actively on board with this plan." For in that case, you have conveyed to your neighbor that you plan to mow his lawn, and that this plan depends not on your whims or desires but on his active endorsement of the plan. If you change your mind and don't mow the lawn in spite of his continued uptake, it seems appropriate for him to resent or forgive you, even if you do warn him in time. Again, the landscaper does not seem similarly positioned to resent or forgive you, since the resolution was not conditional on her agreement. Your resolution yields not just a generic obligation owed to anyone who overheard you, but an obligation that is directed specifically towards your neighbor. In general, if you resolve to Φ conditional on the agreement, acceptance, or uptake of A, it seems that you owe a directed obligation to A to Φ. Promissory resolutions are always directed in this sort of way, because they are always conditional on the acceptance of the promisee.

Third, conditional resolutions generate moral obligations because resolving conditionally on A's acceptance and then failing to act when this condition is met fails to take proper consideration of A's interests. To fail to adequately account for another person's interests when you are engaged in a direct interaction with them is an interpersonal, moral sort of failing. It betrays the wrong kind of attitude to take towards another person, and can be construed as a form of disrespect or a problematic lack of moral concern. For example, suppose I proclaim to my family that I've resolved to give a particular heirloom to my cousin A, so long as she agrees to take it. She agrees, but I change my mind and give it to a different cousin, B. Assume A has no independent claim over the heirloom that B lacks; had I not formed a conditional resolution, it would be permissible for me to give the heirloom to A or B. Since I expressed a conditional resolution to give the heirloom to A, I seem to be *slighting* or *wronging* A when I give the heirloom to B instead. For I am failing to take her interests into account as I should and to give them their proper weight. My expressing a resolution conditional on A's interests entails that she must be given special consideration; in the case where I'm simply deciding between A and B without expressing any resolution, I don't wrong or slight A by choosing

has rearranged her entire schedule because she thought she didn't need to mow your neighbor's lawn. But generally, expectation-based obligations can be satisfied either by acting or altering the expectations.

B, even if A wants the heirloom more than B does. Promissory resolutions are those in which you are morally required to take special consideration of the promisee's interests. To fail to do so is to fail to properly respect the promisee, just as failing to give the heirloom to A when she has accepted my conditional resolution fails to properly respect her. Granted, this may not be a very serious or significant moral failing. But as we saw above, MSF is burdened with establishing the existence only of a relatively weak sort of moral obligation.

In the last two sections, I have offered arguments for two of the core claims that the argument for MSF relies on: (1) that promises express resolutions; (2) that resolutions rationally and morally commit you to acting, all else being equal. These claims are interesting in themselves, as they help us understand the relationship between promises and resolutions, as well as the nature of the commitments incurred by resolutions. With the addition of a third claim—that is, the determination claim—this relationship between promises and resolutions becomes significantly more interesting, for it can generate a new kind of theory of promissory obligation.

At present, I cannot fully defend the determination claim or offer a deep explanation of why it holds. However, I can offer evidence to suggest that the determination claim is both plausible and theoretically fruitful, insofar is it can help explain the norms on speech acts other than promising.

5 Defending Claim (3): Evidence for the Determination Claim

The third claim on which MSF depends is the determination claim, which states that the norms governing the mental state expressed by a speech act (at least partially) determine what the norms on that speech act are, in both sincere and insincere cases. In other words, the determination claim tells us that saying that you're in state X commits you to behaving as if you are in that state, regardless of whether you in fact are.

My first piece of evidence for the determination claim is the observation that, in general, we need to assume the truth of something like the determination claim in order to have fair and productive social interactions. The determination claim states that you must act as if you really are in a particular mental state when you convey to others that you are in that state. People are not mind-readers, and our conversations and social interactions are generally

doi: 10.48106/dial.v74.i4.03

presumed to be cooperative. In light of this, it would be unfair to expect people to be able to ascertain when our utterances are sincere and when they are not. It follows that we must be able to take what people say at face value if we are to have productive interactions with them, at least in typical circumstances.[26] Doing so enables us to respond appropriately to them: to predict what they might do and say next, and to alter our behavior in light of theirs, etc.

If we weren't entitled to presume that people really were in the states they purport to be in, we wouldn't be able to interact with them very effectively. If I promise to Φ, my promisee must be entitled to presume that I am committed to Φing, lest she be at a loss for how to respond to me. And because she cannot know my inner mental state, it would be unreasonable and unfair to expect anything else of her. Accordingly, my promisee is entitled to interact with me as if I have in fact resolved to Φ—which might involve her believing that I plan to Φ, structuring her future plans and behavior around my Φing, or simply responding appropriately in the moment to my commitment to Φing. And as promisor, I should in turn behave as if I really have resolved, and thereby enable my interlocutor to take my promissory utterance at face value—which is to say, I should act in accordance with the determination claim.

The rest of my evidence for the determination claim is circumstantial: there exist other cases in which publicly conveying that you are in a mental state M by performing a speech act S commits you to acting in whatever way is required by M, regardless of whether you really are in M. Promising is not unique in this regard, but is part of a general pattern. This gives us good reason to think that the determination claim is broadly true, and is not merely an *ad hoc* principle that applies only to the case of promises and that I am invoking out of the blue to defend MSF.

26 Because the mental state is explanatorily prior to the speech act (see footnote 13), the mental state is also explanatorily prior to the action that results from the speech act. Suppose that I am in state X, and I convey this by saying "I'm in X." How others can reasonably expect me to act on the basis of this utterance is fundamentally determined not by my utterance that I'm in X, but by what *being in state X* commits me to. To be fair to others, I must act as if I am in state X, unless there is some reason not to take my utterance at face value—say, if it is mutually understood that I am in a strategic (e.g. game-playing) context.

5.1 *Forgiving*

It is plausible that forgiving someone expresses that you have repudiated or foresworn blaming them.[27] Repudiating blame of A for doing X plausibly commits one to ceasing blaming A for X in the moment, and to refraining from actively expressing blame towards A for X again in the future. Publicly expressing forgiveness—regardless of whether you have privately foresworn blame—likewise seems to commit you to refraining from expressing blame in these same ways. That is, publicly conveying that you have repudiated blame by performing a speech act of forgiveness commits you to behaving as if you really have repudiated blame.

To illustrate, suppose that Anna forgets that today is her wedding anniversary, and she fails to meet her spouse Betty for a celebratory dinner they have planned. Betty knows that she is likely to hold this mistake against Anna, and doesn't want to damage their relationship by doing so. So she decides to foreswear blaming Anna for her oversight. This private foreswearing of blame commits her to refraining from expressing blame towards Anna. And telling Anna that she has forgiven her likewise commits Betty to refraining from expressing blame towards Anna. This is so even if Betty is insincere, and forgives Anna not because she has in fact foresworn blaming her, but because she wants to avoid conflict. Betty's interpersonal utterance of forgiveness nevertheless commits her to refraining from openly expressing blame. As with promising, the public expression of a foreswearing of blame changes the nature of the norms to which you are subject; it transforms a private commitment to refraining from engaging in blaming activities into an interpersonal, directed obligation to avoid continuing to blame the wrongdoer in the future. Finally, and again as with promising, this is compatible with the existence of additional explanations of why you should refrain from expressing blame in particular cases; perhaps Betty has an obligation to refrain from expressing blame because she has publicly expressed that she has forgiven Alice, *and* because she is independently obligated to avoid acting unfairly, and expressing blame would be unfair since Betty forgot their anniversary last year.

27 This is a common view of forgiveness in both everyday practice and the philosophical literature; the view is usually attributed first to Bishop Butler. See discussion of this view in Griswold (2007).

doi: 10.48106/dial.v74.i4.03

5.2 *Apologizing*

Similarly, apologizing for Φing expresses that you regret or are sorry that you Φed. When someone regrets an action in this way, they incur a commitment to actively taking responsibility for it somehow. What this involves will vary in different cases; a cheating spouse who regrets their infidelity is obligated to avoid straying again, while a party guest who regrets spilling red wine on a white rug incurs a commitment to clean up the spill. Philosophical and popular consensus is that someone who publicly apologizes for Φing likewise commits themselves to taking responsibility for their action.[28] And this is so regardless of whether the apology was sincere; publicly conveying that you regret your action by apologizing commits you to behaving as if you really have regretted your action.

For example, suppose Christa catches her student Danny using his phone during class, in violation of her policy. Danny apologizes to Christa, which commits him to taking responsibility for his error (e.g. by admitting that he was wrong and refraining from using his phone in class again). This is so even if Danny is insincere, and apologizes only because he fears that Christa will dock his participation grade if he does not. As with promising and forgiveness, a public apology transforms a personal feeling of regret that privately commits you to making amends into an interpersonal demand to make such amends. There may also be additional moral considerations present that require agents to take active responsibility for their wrongdoing. But even if these considerations are not present, the mere (sincere or insincere) expression of regret commits you—at least in a weak way—to taking responsibility for your wrongdoing.

5.3 *Asserting*

There is much disagreement about what mental state assertion expresses and whether there are norms of further commitment on assertions; adjudicating between these views is too large a project to be adequately handled here. But we can assume a particular view to illustrate how assertion might pattern

28 For example, Mihaela Mihai (2013) notes that while philosophical accounts of apology vary, "there is a growing consensus that an authentic apology implies an acknowledgement that the incident in question did in fact occur and that it was inappropriate, a recognition of responsibility for the act, the expression of an attitude of regret and a feeling of remorse, and the declaration of an intention to refrain from similar acts in the future."

with promising, forgiving, and apologizing. Suppose for the sake of argument that assertion expresses belief (which is compatible with assertion expressing something else that includes belief as a component, such as justified belief or knowledge). If you believe that *p* and are questioned about whether *p*, it is plausible that you are normatively committed to defending or justifying *p*. It is also plausible that this same commitment is inherited by assertions: some philosophers argue that someone who asserts that *p* takes on a special commitment to the truth of *p*, which can be cashed out as a commitment to justify or defend p to one's interlocutors. [29] This is plausibly so even in insincere cases; if you are not going to retract your insincere assertion that *p*, you should be prepared to defend or justify it. If this is the case, then assertion patterns with the other speech act/mental state pairs we've been discussing: publicly conveying that you believe *p* by asserting that *p* commits you to behaving as if you really believe *p*, and transforms your private commitment to behave as if *p* is true into a commitment to defending *p* publicly.

5.4 *Conclusion*

The determination claim is a natural and plausible explanation of the pattern outlined above, for two reasons. First, speech acts are dependent on the mental states they express, in a way that mental states are not dependent on the speech acts used to express them. You can properly and without any insincerity be in a particular mental state without expressing it via a speech act; it is no problem to resolve to act without promising that you will do so. But all of the speech acts we have been discussing express that the agent is in a particular mental state, and it is communicatively insincere to perform the speech act without being in that mental state. The direction of explanation proposed by the determination claim tracks this dependency; it would not make sense for the direction of explanation to go in the other direction.

Because we have reason to think the determination claim is true—and we also have reason to think that promises express resolutions, and that resolutions rationally (and sometimes morally) commit us to acting—we have reason to think that MSF is a viable theory of promissory obligation. And MSF has appealing explanatory advantages. Since the mental state expressed by a promise is both distinctive of promises and present in all cases, MSF

29 See, among others, Peirce (1935), Searle (1969), Brandom (1983), Wright (1992), Watson (2004), and MacFarlane (2005).

doi: 10.48106/dial.v74.i4.03

captures the minimal, essentially promissory obligation that is always there, while being open to the pluralist idea that other theories can explain why we have stronger moral reasons to keep our promises in many cases. Moreover, the success of MSF should lead us to be optimistic about the possibility of providing similarly structured and equally resourceful accounts cashing out the norms on other sorts of speech acts in terms of their underlying mental states.*

Alida Liberman
Southern Methodist University
aliberman@smu.edu

References

BRANDOM, Robert B. 1983. "Asserting." *Noûs* 17: 637–50. doi:10.2307/2215086.

BRATMAN, Michael E. 1987. *Intentions, Plans and Practical Reason*. Cambridge, Massachusetts: Harvard University Press.

BROOME, John A. 1999. "Normative Requirements." *Ratio* 12 (4): 398–419. doi:10.1111/1467-9329.00101.

CHOLBI, Michael J. 2002. "A Contractualist Account of Promising." *The Southern Journal of Philosophy* 40 (4): 475–91. doi:10.1111/j.2041-6962.2002.tb01913.x.

COTTINGHAM, John G. 1985. "Review of Robins (1984)." *The Philosophical Quarterly* 35 (140): 315–18. doi:10.2307/2218913.

DOWNIE, R. S. 1985. "Three Accounts of Promising." *The Philosophical Quarterly* 35 (140): 259–71. doi:10.2307/2218905.

GRISWOLD, Charles L. 2007. *Forgiveness. A Philosophical Exploration*. Cambridge: Cambridge University Press.

HEUER, Ulrike. 2012. "Promising – Part 1." *Philosophy Compass* 7 (12): 832–41. doi:10.1111/j.1747-9991.2012.00524.x.

HOLTON, Richard. 2009. *Willing, Wanting, Waiting*. Oxford: Oxford University Press.

HOOKER, Brad. 2011. "Promises and Rule-Consequentialism." In *Promises and Agreements: Philosophical Essays*, 235–52. Oxford: Oxford University Press.

HUME, David. 1739. *A Treatise of Human Nature*. London: John Noon, at the White-Hart.

* This paper is adapted from my dissertation, *The Mental States Theory of Promising,* completed at the University of Southern California. More people than I can list here have helped me to articulate and refine these ideas over the years. Special thanks are due to my committee members Steve Finlay, Jake Ross, and Gary Watson, to my graduate school classmates (especially Josh Crabill, N.G. Laskowski, Michael Milona, Caleb Perl, and Abelard Podgorski), to audiences at USC and Bowling Green State University, and most importantly to my dissertation supervisor Mark Schroeder.

KOLODNY, Niko. 2005. "Why Be Rational?" *Mind* 114 (455): 509–63. doi:10.1093/mind/fzi509.

KOLODNY, Niko, and Richard Jay WALLACE. 2003. "Promises and Practices Revisited." *Philosophy & Public Affairs* 31 (2): 119–54. doi:10.1111/j.1088-4963.2003.00119.x.

LEMOS, Noah M. 1987. "Review of Robins (1984)." *Philosophy and Phenomenological Research* 47 (4): 685–88. doi:10.2307/2107242.

LIBERMAN, Alida. 2015. "The Mental States First Theory of Promising." PhD dissertation, University of Southern California.

———. 2016. "Reconsidering Resolutions." *Journal of Ethics and Social Philosophy* 10 (2): 1–27. doi:10.26556/jesp.v10i2.98.

———. 2020. "Consequentialism and Promises." In *The Oxford Handbook of Consequentialism*, edited by Douglas W. Portmore, 289–309. Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780190905323.013.16.

LIBERMAN, Alida, and Mark SCHROEDER. 2016. "Commitment: Worth the Weight." In *Weighing Reasons*, edited by Errol Lord and Barry Maguire, 104–20. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199315192.003.0005.

MACFARLANE, John. 2005. "Making Sense of Relative Truth." *Proceedings of the Aristotelian Society* 105 (3): 321–61. doi:10.1111/j.0066-7373.2004.00116.x.

MARUŠIĆ, Berislav. 2013. "Promising Against the Evidence." *Ethics* 123 (2): 292–317. doi:10.1086/668704.

MIHAI, Mihaela. 2013. "Apology." In *Internet Encyclopedia of Philosophy*. https://iep.utm.edu/apology/.

OWENS, David. 2012. *Shaping the Normative Landscape*. Oxford: Oxford University Press.

PEETZ, Vera. 1977. "Promises and Threats." *Mind* 86 (344): 578–81. doi:10.1093/mind/LXXXVI.344.578.

PEIRCE, Charles Sanders. 1935. *Collected Papers of Charles Sanders Peirce: Pragmatism and Pragmaticism*. Edited by Charles Hartshorne and Paul Weiss. Vol. 5. Cambridge, Massachusetts: Harvard University Press.

PINK, Thomas. 2009. "Promising and Obligation." In *Philosophical Perspectives 23: Ethics*, edited by John Hawthorne, 389–420. Oxford: John Wiley; Sons, Inc.

RAWLS, John. 1971. *A Theory of Justice*. Cambridge, Massachusetts: Harvard University Press.

RAZ, Joseph. 1977. "Promises and Obligations." In *Law, Morality and Society. Essays in Honour of h.l.a. Hart*, edited by Peter M. S. Hacker and Joseph Raz, 210–28. Oxford: Oxford University Press.

ROBINS, Michael H. 1984. *Promising, Intending and Moral Autonomy*. Cambridge Studies in Philosophy. Cambridge: Cambridge University Press.

SCANLON, Thomas Michael. 1990. "Promises and Practices." *Philosophy and Public Affairs* 19 (3): 199–226.

———. 1998. *What We Owe to Each Other*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

SEARLE, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

SHEINMAN, Hanoch. 2008. "Promise as Practice Reason." *Acta Analytica* 23 (4): 287–318. doi:10.1007/s12136-008-0033-1.

SHIFFRIN, Seana Valentine. 2008. "Promising, Intimate Relationships, and Conventionalism." *The Philosophical Review* 117 (4): 481–524. doi:10.1215/00318108-2008-014.

SHPALL, Samuel. 2013. "Wide and Narrow Scope." *Philosophical Studies* 163 (3): 717–36. doi:10.1007/s11098-011-9841-z.

———. 2014. "Moral and Rational Commitment." *Philosophy and Phenomenological Research* 88 (1): 146–72. doi:10.1111/j.1933-1592.2012.00618.x.

SMITH, Holly M. 1987. "Review of Robins (1984)." *Noûs* 21 (4): 604–8. doi:10.2307/2215676.

SOUTHWOOD, Nicholas, and Daniel FRIEDRICH. 2009. "Promises Beyond Assurance." *Philosophical Studies* 144 (2): 261–80. doi:10.1007/s11098-008-9209-1.

THOMSON, Judith Jarvis. 1990. *The Realm of Rights*. Cambridge, Massachusetts: Harvard University Press.

WATSON, Gary. 2004. "Asserting and Promising." *Philosophical Studies* 117 (1–2): 57–77. doi:10.1023/B:PHIL.0000014525.93335.9e.

WAY, Jonathan. 2010. "The Normativity of Rationality." *Philosophy Compass* 5 (12): 1057–68. doi:10.1111/j.1747-9991.2010.00357.x.

WRIGHT, Crispin. 1992. *Truth and Objectivity*. Cambridge, Massachusetts: Harvard University Press.