

ARTICLES

THE RATIONALITY OF PROMISING

*Emily Sherwin**

INTRODUCTION	533
I. SOME THEORIES OF PROMISING	534
II. PRACTICAL RATIONALITY	538
III. EPISTEMIC RATIONALITY	540
IV. PRACTICAL AND EPISTEMIC CIRCUMSTANCES OF PROMISORS	541
V. CONFLICTING STANDARDS OF RATIONALITY	545
A. <i>The Role of Beliefs in Extended Practical Rationality</i>	545
B. <i>The Ethics of Belief</i>	546
CONCLUSION	551

INTRODUCTION

Binding promises yield a number of practical benefits, if in fact they are binding. One benefit is coordination. Knowing that she must perform, the promisor can allocate her time and resources more effectively. The promisee, meanwhile, can make plans on the assumption that the promised act will occur.

Markets for future exchange rely on the coordinating power of binding promises. For this purpose, it may be possible in theory to support coordination by designing and enforcing an ideal set of legal rules governing contractual obligation. Almost certainly, however, markets will function more effectively if promises also impose obligations to perform on those who make them.

Apart from coordination, binding promises allow parties to alter the normative relations between them. By voluntarily communicating an intention to be bound, the promisor assumes an obligation to the promisee.¹ In this way, the promisor can both control her own future action and

* Frank B. Ingersoll Professor, Cornell Law School.

¹ See Joseph Raz, *Is There a Reason to Keep a Promise?*, in *PHILOSOPHICAL FOUNDATIONS OF CONTRACT LAW* 58, 60 (Gregory Klass, George Letsas, & Prince Saprui, eds., 2014) [hereinafter Raz, *Is There a Reason?*]; Joseph Raz, *Promises and Obligations*, in *LAW, MORALITY, AND SOCIETY* 210, 211 (P.M.S. Hacker & J. Raz eds., 1977) (proposing an obligation-based account of promising, as opposed to an intention-based account, and discussing the normative functions of promises) [hereinafter Raz, *Promises and Obligations*].

confer power on the promisee either to demand what was promised or to release the promisor.² At least, the promisor can change the normative picture in this way if the promise is in fact binding.

I begin this Article by canvassing some competing conceptions of promissory obligation, which point to a range of benefits associated with promising. I then turn to theories of practical rationality, showing how it might, under theories of temporally extended practical rationality, be practically rational to treat a promise as binding. I then take a brief look at standards of epistemic rationality. Next, I consider the epistemic circumstances of a promisor, both at the time of promising and at the time of performance. I conclude that despite whatever practical and normative benefits binding promises may have, it will not always be epistemically rational for promisors to perform. To the extent that markets rely on binding promises, therefore, they rely on an element of irrationality in human decision-making.

I. SOME THEORIES OF PROMISING

There is no consensus on the question whether promises and similar forms of commitment in fact alter normative relations between parties, or, if they do, on how exactly the alteration comes about. Hume took the position that there is no mental act capable of generating a moral obligation to act.³ Promises become binding, if at all, by virtue of social practices designed to advance social interests, which lead others to expect performance.⁴ According to views of this type, any normative responsibilities the promisor incurs by promising depend on the promisor's pre-existing obligation not to do harm rather than a self-imposed duty to perform.

T.M. Scanlon offers a theory of promissory obligation that does not depend on social practice, but also denies the possibility of self-imposed obligation.⁵ Scanlon argues that when one person leads another to believe that she will take some future action, knowing that assurance of future action is important to the other, she becomes obligated to perform

² See Raz, *Is There a Reason?*, *supra* note 1, at 61. In Hohfeldian terms, the promisee has a right to performance (or compensation for failure to perform) and a power to eliminate the obligation to perform. See Wesley N. Hohfeld, *Some Fundamental Conceptions as Applied in Judicial Reasoning*, 23 *YALE L.J.* 16, 29–32 (1913). In Hohfeldian terms, the promisor has a right to performance (or compensation for failure to perform) and the power to eliminate the obligation to perform.

³ See DAVID HUME, *A TREATISE OF HUMAN NATURE* 516 (L.A. Selby & P.H. Nidditch eds., 2d ed. 1978).

⁴ See *id.* at 516–17. See also John Rawls, *Two Concepts of Rules*, 64 *PHIL. REV.* 3, 17 (1955) (describing the bindingness of promises as a rule imbedded in a beneficial practice and thus impervious to general utilitarian balancing).

⁵ See THOMAS SCANLON, *WHAT WE OWE TO EACH OTHER* 295–317 (1998).

as expected.⁶ Because the promisee's interest is in assurance that the promised act will occur, the duty that results is not just a duty to compensate for harm but a duty to perform.⁷ The duty Scanlon describes is not willed into existence by the promisor, but depends instead on a set of background normative obligations that one person owes to another when her actions generate expectations.⁸

A different line of argument holds that autonomous individuals have the power to impose new, binding obligations on themselves through an act of will. Charles Fried, for example, has developed a theory of contract law that links promissory obligations to "a morality of autonomy, respect for persons, and trust."⁹ David Owens argues that rights and duties are important human goods, which we have an interest in controlling through mechanisms such as promising.¹⁰ Our interest in control supports the existence of a normative power to stipulate that failure to perform a promised act will count as a wrong.¹¹ Similarly, Joseph Raz suggests that in making a promise, the promisor confers on the promisee a valuable power to control the promisor's future actions, and in doing so incurs an obligation to perform.¹²

Another influential theory of promissory obligation comes from Seana Shiffrin, who proposes that promising plays a key role in maintaining "morally decent" human relationships, which in turn are fundamental to human autonomy.¹³ The power to promise allows the parties to a relationship to make assurances, secure trust, settle differences, and right imbalances of power.¹⁴ Because we presumptively are capable of morally sound relationships, and because morally sound relationships depend on the power to obligate ourselves, we must possess this normative power.¹⁵

⁶ *Id.* at 296–97.

⁷ *Id.*

⁸ *Id.*

⁹ See Charles Fried, *The Ambitions of Contract as Promise*, in *PHILOSOPHICAL FOUNDATIONS OF CONTRACT LAW*, *supra* note 1, at 17, 20. See generally Charles Fried, *CONTRACT AS PROMISE: A THEORY OF CONTRACTUAL OBLIGATION* (1981) (setting out the theory).

¹⁰ David Owens, *Does a Promise Transfer a Right?*, in *PHILOSOPHICAL FOUNDATIONS OF CONTRACT LAW*, *supra* note 1, at 78, 79–81.

¹¹ *Id.*

¹² For Raz, the obligation that results is normative in the sense that it creates a new reason for action. See Raz, *Is There a Reason to Keep a Promise?*, in *PHILOSOPHICAL FOUNDATIONS OF CONTRACT LAW*, *supra* note 1, at 58, 60–61. Whether a promissory obligation is a moral one depends on "the content and circumstances of [the] particular promise" rather than the fact that the promisor has made a binding promise. Raz, *Promises and Obligation*, in *LAW, MORALITY, AND SOCIETY*, *supra* note 1, at 210, 225.

¹³ Seana Shiffrin, *Promising, Intimate Relationships, and Conventionalism*, 114 *PHIL. REV.* 481, 499 (2008).

¹⁴ *Id.* at 497–510, 518–19.

¹⁵ *Id.*

The initial problem with theories of promissory obligation that rely on autonomy and other values that might be served by the power to impose new obligations on oneself is that the potential benefits of having such a power do not establish that it exists.¹⁶ A further difficulty is that, assuming the value of normative self-control supports the power to promise, it does not follow that the value of normative self-control supports an obligation to perform the promise. As Raz points out, keeping a promise may avoid a wrong, but it does not add to the promisor's self-control. A promisor who breaks a binding promise does not lose the power to promise and her later promises have the same moral significance as they would have had if she had kept the first promise. Accordingly, the value the promisor places on self-control cannot provide a full explanation for the binding of promises.¹⁷

Shiffrin offers several arguments for the binding of promises.¹⁸ One of these arguments explicitly endorses the idea suggested by Raz, that the promisor transfers something of value to the promisee.¹⁹ Specifically, the promisor gives the promisee a right to decide whether the promisor should perform the promised act.²⁰ This explanation connects the power to promise to the duty to perform and also connects the duty to perform to the time of the promise, but it raises other difficulties.²¹ The nature of transferred rights is not evident, and the rights the promisee receives may not perfectly correspond to the rights the promisor gives up.²²

¹⁶ Raz acknowledges this difficulty, saying that “[i]t is impossible to have the power to promise, however good it may be to have it, unless [the fact] that one promised is a reason to do as one promised.” Raz, *Is There a Reason?*, *supra* note 1, at 69. “Yet,” he goes on to say, “we do have it.” *Id.*

¹⁷ *Id.* at 67–69. Raz’s account has the virtue of tying the reasons for action created by a promise to performance of the promised act. Yet it does not fully explain how the promisor is able to give the promisee a valuable assurance of performance at the time of the promise. The promisor is obligated to perform at T_2 if the promisee gives valuable assurance to the promisee at T_1 . The assurance is valuable to the promisee if the promisor will be obligated at T_2 to perform. The promisor will be obligated at T_2 to perform if the promisor gave valuable assurance of performance at T_1 , and so forth. There is no obvious way to escape this loop of conditions.

¹⁸ Shiffrin, *supra* note 15, at 517–19.

¹⁹ *Id.* at 517. *But cf.* David Owens, *Does a Promise Transfer a Right?*, in *PHILOSOPHICAL FOUNDATIONS OF CONTRACT LAW*, *supra* note 1, at 78 (2012) (rejecting transfer theories).

²⁰ Shiffrin, *supra* note 13, at 517. Kant suggested a transfer theory of promising, stating that “. . . what I acquire directly by a contract is not an external thing but rather his deed, by which that thing is brought under my control so that I make it mine”. IMMANUEL KANT, *THE METAPHYSICS OF MORALS* § 20, at 93 (Mary J. Gregor trans., Cambridge Univ. Press 1991) (1797). *See also* Andrew S. Gold, *A Property Theory of Contract*, 10 *NW.U. L. REV.* 1, 2 (2009) (proposing a transfer theory of contractual obligation limited to cases of fair mutual exchange).

²¹ *But see* Gold, *supra* note 20, at 2 (connecting the duty to perform to the future time of performance and not the time of the promise).

²² *See* Stephen A. Smith, *Toward a Theory of Contract*, in *OXFORD ESSAYS IN JURISPRUDENCE* 107, 120 (J. Hurder ed., OUP 2000).

Another argument Shiffrin makes for the bindingness of promises at the time of performance is that promises, particularly within close relationships, solicit the promisee's trust.²³ By acting inconsistently with her solicitation of trust, the promisor commits a wrong against the promisee.²⁴ This argument, however, resembles Scanlon's argument, which is not an argument for normative power.²⁵ Consequently, it fails to explain Shiffrin's initial assertion that the promisor binds herself by an act of will rather than by subjecting herself to pre-existing normative constraints against harm.

Assuming it is possible for an individual to provide herself with a new reason for action by intentionally undertaking an obligation, a further question is what weight such a reason carries. The reason generated by a promise is sometimes referred to as a content-independent reason for action, meaning that the reason reflects the promise itself, independent of what was promised and what consequences a breach of the promise may cause.²⁶ Raz, however, points out that a content-independent reason risks being a bare reason, meaning a reason that has no weight when balanced against contrary reasons for action.²⁷ A bare reason not only lacks weight but also cannot generate secondary reasons for action such as promisee reliance or the negative effect of a breach on the general reliability of promises.²⁸ Secondary reasons of this kind depend on the presupposition that the promise itself creates some initial reason to perform.²⁹ If there is nothing that could, in principle, determine the weight of the initial promissory reason, then the reason is not a reason at all and the secondary reasons that depend on it do not come into play.³⁰

Raz's answer to this problem is that the weight of the initial promissory reason for action generated by a promise is determined by the value of the assurance the promise provides, in the context of each particular promise.³¹ Even if the promisee does not value the promised act, promissory assurance can be valuable because it provides the promisee with an opportunity to become interested in performance, or because it gives the promisee the power to choose whether or not to demand performance.³² Accordingly, Raz concludes that promissory reasons are not bare rea-

²³ See Shiffrin, *supra* note 13, at 496–97.

²⁴ See *id.* at 486–87.

²⁵ See *id.*; Scanlon, *supra* note 5, at 295–96.

²⁶ H.L.A. Hart referred to reasons arising from promising and other forms of commitment as “content-independent” reasons for action. See H.L.A. HART, *ESSAYS ON BENTHAM* 254–55 (1982).

²⁷ Raz, *Is There a Reason?*, *supra* note 1, at 63.

²⁸ *Id.* at 63–65. See *id.* at 67–71.

²⁹ *Id.* at 63–65.

³⁰ Raz, *Is There a Reason?*, *supra* note 1, at 71.

³¹ Raz, *Is There a Reason?*, *supra* note 1, at 76–77.

³² *Id.*

sons, but reasons whose weight is ascertainable in principle, if not in practice.³³

Reviewing the possibilities: some reject the claim that promises generate obligations independent of the harm that may follow from a failure to perform. Others maintain that agents have normative power to impose obligations on themselves by choosing to make a promise. Those who endorse the possibility of an independent self-imposed promissory obligation may disagree about how the obligation arises, whether it counts as a moral obligation, and what weight it carries when balanced against other considerations.

II. PRACTICAL RATIONALITY

Practical rationality is the rationality of actions, or of agents in choosing actions. Practical rationality is most often described in instrumental terms: an action is rational if it is effective in serving the agent's ends and an agent is rational if the processes by which she decides upon action is likely to produce instrumentally rational actions.³⁴ More substantive standards can be incorporated by folding them into a notion of rational ends.

Traditional standards of practical rationality refer to the agent's current reasons for action at the time she acts. A number of writers, however, have proposed that practical rationality should be understood as extending over time, to allow for advance planning. David Gauthier, Edward McClennen, and Scott Shapiro all have proposed that, to realize the instrumental benefits of planning, agents must to some degree be bound by their prior intentions. Gauthier suggests that if the agent adopted a course of action expecting that it will be instrumentally beneficial, and continues to believe at the time of action that the overall course of action is beneficial, the course of action takes priority over current reasons not to perform the particular act that completes it.³⁵ McClennen proposes that a rational agent should treat her prior choices as "resolute," at least if doing so will yield a net benefit to all the agent's past and future selves,

³³ See *id.* at 71.

³⁴ See, e.g., MICHAEL E. BRATMAN, INTENTIONS, PLANS, AND PRACTICAL REASON 50–53 (1987) (discussing instrumental rationality and agent rationality).

³⁵ David Gauthier, *Intention and Deliberation*, in MODELING RATIONALITY, MORALITY, AND EVOLUTION, 41, 44, 48–49 (P. Danielson ed., 1998) (discussing the rationality of following intentions).

fairly distributed among them.³⁶ Shapiro argues that an agent who has formed an intention to act has no rational option but to act as intended.³⁷

Although Gauthier, McClennen, and Shapiro are correct that synchronic judgments about reasons for action are incomplete, I set their versions of practical rationality to one side for several related reasons. First, none of these theories provide a cognitive explanation of how prior intentions can constrain an agent's actions when the agent has in fact formed a contrary belief about her current reasons for action. Second, it seems likely that, as Joseph Raz and others have argued, an intention cannot lead directly to action without the additional element of will to act.³⁸ An agent who believes that her current reasons for action, fully considered, do not support acting on her intention normally will lack the will to carry through.

Consequently, I will focus on the account of temporally extended practical rationality provided by Michael Bratman. Summarizing a complex theory very briefly: Bratman's approach to practical rationality relies on a special understanding of the nature and function of intentions.³⁹ Bratman views intentions as independent cognitive states that are not simply composites of beliefs and desires. They play a special cognitive role in the process leading to action because as long as the intention remains in place, it carries with it two forms of commitment, which Bratman calls "volitional" commitment and "reason-centered" commitment.⁴⁰ Volitional commitment means that as long as the intention remains in place, it generates a "pro-attitude," capable of supporting action without new deliberation. Same-centered commitment means that, while in place, the intention limits the options open to the agent in deliberation about further action or intentions. If the agent reconsiders her intention

³⁶ See generally EDWARD F. MCCLENNEN, RATIONALITY AND DYNAMIC CHOICE 12–13, 120–22 (CUP 1990) (discussing rational choice); Edward F. McClennen, *Pragmatic Rationality and Rules* 26 PHIL. & PUBL. AFFAIRS 210, 232 (1997) (discussing rational choice); Edward F. McClennen & Scott J. Shapiro, *Rule-Guided Behavior*, in III NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW 367 (Peter Newman ed., Stockton Press 1998) (discussing rational choice).

³⁷ SCOTT J. SHAPIRO, THE DIFFERENCE THAT RULES MAKE, in ANALYZING LAW: NEW ESSAYS IN LEGAL THEORY 33, 39–40, 47 (Brian Bix ed., Clarendon Press 1998) (discussing binding intentions).

³⁸ See JOSEPH RAZ, FROM NORMATIVITY AND RESPONSIBILITY 131–32 (2011); John Broome, *Practical Reasoning*, in REASON AND NATURE: ESSAYS IN THE THEORY OF RATIONALITY (J. Bermúdez & Alan Millar eds., 2002). Possibly McClennen could be read as addressing this problem. McClennen requires as a condition of extended practical rationality that acting on a prior intention must serve the collective interest of all the agent's selves over time and that benefits must be fairly distributed among selves. McClennen & Shapiro, *supra* note 36, at 367. It might follow that if the agent, as now constituted, is viewed as embodying all her past and present selves, she could will action on behalf of the group. Will, however, is most plausibly understood as the will of the agent's current self alone.

³⁹ See generally BRATMAN, *supra* note 34 (presenting the planning theory of intention).

⁴⁰ *Id.* at 15–18.

before acting on it, her prior intention dissolves and the volitional commitment that accompanied it is gone. If she does not reconsider or block her intention, she proceeds to act without further reflection on reasons for action.

The standards Bratman proposes to govern the practical rationality of an agent who acts unreflectively on prior intentions are designed to achieve reasonable, although not ideal, stability of intentions. The agent is rational if her intention is rational when formed and if she develops and follows a reasonable set of dispositions toward retaining or reconsidering prior intentions. The agent's dispositions toward prior intentions are reasonable if following them will produce better results over the long run than judging what to do at the time of action. Thus, for an agent who rationally forms the intention to make and perform a promise, and is reasonably disposed to keep her promises in most or all cases, and does not reconsider her intention before the time of performance, it is practically rational to honor the promise.

III. EPISTEMIC RATIONALITY

Epistemic rationality is the rationality of belief, or of agents in forming beliefs.⁴¹ Theories of epistemic rationality vary along many dimensions, although most require some connection between the grounds of an agent's beliefs and their probable truth. For the purpose of this Article, I will make a number of assumptions about epistemic rationality that are fairly widely accepted. I assume an evidential standard of justification for belief, which makes justification depend on the fit between the agent's beliefs and the evidence in her possession.⁴² I assume that epistemic rationality is synchronic: at any point in time, the question is whether the agent's beliefs match her current evidence. I also assume that, although epistemic rationality is judged from an external point of view, standards of epistemic rationality measure the agent's beliefs against *her* evidence, and require a reasonable, rather than an ideal, fit between evidence and belief.⁴³ The agent is not required to draw flawless

⁴¹ RICHARD FOLEY, *THE THEORY OF EPISTEMIC RATIONALITY* (1987).

⁴² The leading alternative to evidentialism is process-reliabilism, which requires the agent to form her belief through processes that generally lead to true beliefs. *See e.g.*, Alvin Goldman, *What is Justified Belief?*, in *JUSTIFICATION AND KNOWLEDGE* 9 (G. Pappas ed., 1979).

⁴³ The standard of reasonableness is somewhat more controversial than other assumptions I have made. *Compare* Earl Conee & Richard Feldman, *Evidentialism*, in *ESSAYS IN EPISTEMOLOGY* 81, 87–89 (2004) (appearing to require a perfect inferential fit), *with* FOLEY, *supra* note 41 at 9–10 (1987) (connecting epistemic rationality to first person persuasive argument under conditions of careful reflection), *and* Hilary Kornblith, *Justified Belief and Epistemically Responsible Action*, 92 *PHIL. REV.* 33, 46 (1983) (arguing that “having justified beliefs is simply doing the best one can in light of the innate endowments one starts from, however reliable or unreliable it may be”).

inferences or to avoid the types of cognitive bias that affect most human reasoners.

In the context of promising, the beliefs in question are beliefs about reasons for action, either at the time the agent makes the promise and forms an intention to honor it or at the time of performance. Thus, standards of practical rationality define what counts as a reason for action, mainly in terms of instrumental benefit. Standards of epistemic rationality match reasons for action to the world at a point in time.

A final question is whether epistemic rationality imposes a responsibility on agents to respond to evidence in their possession and form beliefs that match it, or requires only that such beliefs as the agent chooses to form must match the agent's evidence. This question is important for the arguments I will make below. To show why, I will first examine the epistemic position of a promisor at the time of the promise and at the time for performance.

IV. PRACTICAL AND EPISTEMIC CIRCUMSTANCES OF PROMISORS

As described earlier, some take the view that promises do not create a new, self-imposed reason for action, although a breach may cause harm of various kinds.⁴⁴ Others take the view that agents have the capacity to impose obligations on themselves voluntarily by promising, which create reasons to perform the promise. To illustrate how these two approaches affect a promisor's deliberation about whether to perform a promised act, I will place them in the context of a simple example. The example suggests that under either approach, it will not always be rational to perform the promise, even if it was rational initially to make the promise.

Assume that at T_1 , P promises to mow Bob's lawn on Saturday in exchange for \$50 and forms an intention to perform her promise. P has a fairly good moral sense and an inclination to do what is right: she is not a "bad man" who seeks to take advantages of social practices such as promising. At T_1 when she forms her intention to perform, she believes that her mowing the lawn and Bob's paying her \$50 will yield net benefits for each of them. She believes that breaching the promise will have a negative effect on the coordination benefits of the practice of promising, which allow people like Bob and S to coordinate their activities more effectively than they could if they guessed what each of them would do on Saturday. Further, she believes that a breach could cause unspecified inconvenience to Bob and harm her own reputation as a reliable lawnmower.

S also believes at T_1 that if she deliberates again on Saturday about what to do, she may err in her assessment. S understands at T_1 that she is

⁴⁴ See Scanlon, *supra* note 5, at 295–96.

not a perfect reasoner. One difficulty is that her evidence at T_2 about the costs and benefits of performing her promise may be incomplete. In particular, she may not understand the full consequences that a breach might cause for Bob or for herself. Another difficulty is that like all human reasoners, S is subject to cognitive biases that may cause her to overvalue the benefits of a breach and undervalue its costs.⁴⁵ In particular, she is susceptible to a common bias in favor of immediate and salient facts, which may cause her to disregard or minimize the harm her breach may cause to the practice of promising. Consequently, S believes at T_1 that she will tend systematically to err in favor of breaching her promise.

At the same time, S understands at T_1 that if she judges on Saturday that, all things considered, she should breach the promise, her judgment may be correct. Circumstances on Saturday are still unknown and it is possible that she will have good reasons to break her promise, notwithstanding possible harm to Bob, her own reputation, and the practice of promising. Nevertheless, S concludes at T_1 that, over the range of possible circumstances that may materialize by Saturday and given her propensity to err in assessing them at the time, she will do better on average if she simply performs her promise without further thought than if she reevaluates on Saturday what she should do. Accordingly, she resolves to treat her promise as a binding personal rule requiring her to perform.⁴⁶

Another way to put this is that S resolves to treat her promise as an exclusionary reason for action.⁴⁷ An exclusionary reason operates in two ways: it serves as a first-order reason to perform the promised act, and also as a second-order reason not to act on some range of reasons that weigh against performing. Thus, the agent must either ignore reasons to breach that are subsumed within the exclusionary reason to perform or resist acting on these reasons.

Sometime later, at T_2 , S is called upon to perform her promise. At this time, she continues to hold all the beliefs that led her to make her promise and resolve to treat it as an exclusionary reason for action, based

⁴⁵ Source needed? For overviews of the problem of cognitive biases and studies documenting the availability heuristic and similar patterns of human reasoning, see *JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES* 163 (Daniel Kahneman, Paul Slovic & Amos Tversky, eds., Cambridge: Cambridge University Press 1982); *HEURISTICS & BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT* (Daniel Kahneman, Paul Slovic & Amos Tversky, eds., Cambridge: Cambridge University Press 2002).

⁴⁶ S's personal rule may be a rule to perform all promises, or a rule to perform this promise in all circumstances. Both are general rules because they cover a range of not-yet-specified future cases.

⁴⁷ Joseph Raz proposed in his early writing on promising that promissory obligations function as exclusionary reasons for action. Raz, *Promises and Obligations*, *supra* note X, at 221–23. He did not, however, consider the problem of epistemic rationality or provide a cognitive explanation of how a rational promisor can give exclusionary effect to a promise at the time of performance.

on the evidence and inferences in play at T_1 . Yet, S also has new evidence that bears on the question whether to perform. Assume, for example, that on Friday night, S's pregnant daughter calls to say she has gone into labor three weeks early and needs S's help. S's daughter lives four hours away, making it impossible for P both to mow Bob's lawn on Saturday and to help her child. Lawnmowers are busy at this time of year, but one can probably be found within a reasonable period of time.

If S reflects on this evidence, S may well conclude that current reasons for action favor breaking her promise. Notice that S can reach this conclusion without revising any of the beliefs that led her to conclude at T_1 that she should make the promise and later perform as promised. S may continue to understand that her evidence could be incomplete and that her current assessment of the evidence could be incorrect due to cognitive bias or other inferential mistakes. She may continue to believe that the practice of promising carries important personal and social benefits that may be impaired if she breaches. But she also will understand, as she did at T_1 , that the rule of performance she has imposed on herself is general in the sense that it will not always yield the best outcome on all possible facts. Because S is not a perfect reasoner, she will do better on average if she always performs her promise than if she always judges case-by-case whether to perform. But in some particular cases, her judgment will be correct and following the promissory rule she has set for herself will lead to the wrong result. Thus, if it appears to S that the sum of evidence in her possession supports the inference that she should not now keep her promise, there is no inconsistency between her set of beliefs formed at T_1 and her new belief that reasons for action favor a breach.

In the circumstances just described, the result may seem obvious: S should break her promise in order to help her child. If, on the other hand, S's daughter calls to say she has a Groupon discount for a spa visit on Saturday, most promisors are likely to recognize that the expected harm to Bob and to the reliability of promises outweighs S's desire to opt for the spa. Yet, there will also be cases near the center, in which S might reasonably conclude that her reasons for action favor a breach, but her conclusion may be wrong: her daughter has a work obligation and begs for help with the children.

At least in this middle zone, S faces a dilemma. Standards of long-term practical rationality hold that, given the imperfections of her reasoning in particular cases, it is best that she follow her rule and keep the promise.⁴⁸ Yet, if her current evidence suggests that she should breach,

⁴⁸ See BRATMAN, *supra* note 34, at 28–32.

standards of epistemic rationality require that her current beliefs about reasons for action should match the evidence.

The second approach to promising assumes that agents can and do impose new obligations on themselves when they make promises, and that those new obligations create content-independent reasons to perform the promises. Assuming the agent's new reason is not a bare reason without weight, the promise alters what otherwise would be the promisor's balance of reasons for or against performance at T_2 . Consequently, it may now be rational to perform the promised act at T_2 , although in the absence of a promise it would not be rational to perform the same act.

Nevertheless, a promissory reason to perform may not be enough to solve the problem of rationality. Suppose that at T_1 S concludes that averaging her expected circumstances on Saturday, she will advance her immediate interests by mowing Bob's lawn in exchange for \$50. She also believes that credible promises yield a number of more abstract benefits, including values associated with the power of self-direction. She understands that circumstances might arise in which she will have reason to break the promise, but mistrusts her own future assessment of reasons for action. In addition to ordinary problems of cognitive bias, she may have difficulty quantifying the value of self-direction. Overall, therefore, she believes that over the range of circumstances that may occur at the time of performance, she will do better on average by always performing her promise than by judging what to do on a case-by-case basis. Accordingly, S elects both to assume an obligation to perform and to treat her obligation as an exclusionary reason to perform at T_2 .⁴⁹

If S's promise generates a conclusive reason to perform, or transfers S's power of choice to Bob, then S will not face a problem at T_2 . By hypothesis, her intention to treat the promise as an exclusionary reason will match her reasons for action. This interpretation of the effect of a promise, however, seems extreme: many promises are fairly trivial, and other obligations may intervene.

If the obligation to perform a promise is less than conclusive, the problem persists. Because S's self-imposed obligation and her accompanying intention to treat her promise as an exclusionary reason will produce the best average outcomes over possible circumstances at T_2 , long-term practical rationality requires her to perform.⁵⁰ Yet, she understands that in some particular cases, it will be better for her to breach. Therefore, if her evidence at T_2 indicates that her reasons favor a breach, and if she adverts to this evidence and forms a belief about what she now

⁴⁹ The exclusion need not be complete: there may be a threshold of contrary reasons beyond which P should set aside her intention and reconsider her promise. Within the excluded range, however, she should ignore contrary reasons for action.

⁵⁰ See BRATMAN, *supra* note 34, at 3–11.

should do, epistemic rationality requires that her belief must match her evidence.

V. CONFLICTING STANDARDS OF RATIONALITY

There are two possible ways to argue that there is no inconsistency between practical rationality and epistemic rationality. The first path is to focus on practical rationality and argue that because practical rationality refers only to action, it is indifferent to what the agent believes about reasons for action at the time she acts. In other words, the agent can form a set of epistemically rational beliefs about current reasons for action and still follow the practically rational course of acting on her prior intention. As I will explain, however, this argument is not available given Bratman's assumptions about reconsideration of prior intentions. It also fails to account for the element of will needed to convert intentions into actions.

The second path is to argue that, even if practical rationality requires the agent to respond to beliefs she has formed about current reasons for action, the agent may not form such a belief. This is a key assumption in Bratman's theory.⁵¹ If the agent is not required to form a belief about current reasons, then a theory of long-term practical rationality can succeed in avoiding conflict with epistemic rationality by simply bypassing belief. This argument for reconciling practical rationality and epistemic rationality is more promising. The question it raises is whether epistemic rationality includes a responsibility to advert to evidence about reasons for action before acting on a prior intention.

A. *The Role of Beliefs in Extended Practical Rationality*

I begin with the question what if any impact a new belief about reasons for action have on the practical rationality of acting on a prior intention. If the practical rationality of acting on an intention is unaffected by the agent's beliefs about reasons for action when the time comes to act, then practical rationality and epistemic rationality can easily be reconciled. The agent can form an epistemically rational belief about her current reasons for action but proceed to act on a practically rational prior intention, although her prior intention does not track current reasons for action.⁵²

⁵¹ *Id.* at 54–59.

⁵² As noted earlier, some proponents of temporally extended practical rationality, notably Gauthier, McClennen, and Shapiro, appear to make this assumption. Each suggests that a prior intention constrains current action, despite what the agent may currently believe about reasons for action. Because these theories fail to provide a convincing account of extended practical rationality of agents, I continue to focus on Michael Bratman's theory of temporally extended practical rationality. See BRATMAN, *supra* note 34, at 53–55.

Within Michael Bratman's theory of extended practical rationality, the short answer to the argument that practical rationality is independent of belief is that agents are free to reconsider their prior intentions.⁵³ When an agent comes to believe that she should not now act as she intended to act, she has implicitly reconsidered her either her intention or its current application and it no longer controls her action. Assuming that following the intention continues to be practically rational over the long term, this means that an epistemically rational belief about reasons for action can undermine extended practical rationality.

Second, it seems likely that, as some have argued, an intention cannot lead directly to action without the additional element of will to act.⁵⁴ An agent who believes that her current reasons for action, fully considered, do not support acting on her intention normally will lack the will to carry through. Thus, even if it remains practically rational to perform a promise, the agent may be unable to do so.

B. *The Ethics of Belief*

Assuming that it is either no longer practically rational or no longer feasible for an agent to perform a promise when she has come to believe that her current reasons for action support breaching it, Bratman might argue that, under his theory of action on prior intentions, the agent will not form such a belief. As long as the agent's practically rational dispositions remain intact, the agent will not advert to evidence about reasons for action, will not reflect about whether to follow R, and will not form a belief.⁵⁵ Further, Bratman might add, there is no epistemic reason why she should consult evidence and reflect about whether to follow a rule. Epistemically, she is free to do what practical rationality requires, which

⁵³ See BRATMAN, *supra* note 34, at 50–53.

⁵⁴ See RAZ, *NORMATIVITY AND RESPONSIBILITY*, *supra* note 38, at 131–37; Broome, *supra* note 38 (2002). Possibly McClennen could be read as addressing this problem. McClennen requires as a condition of extended practical rationality that acting on a prior intention must serve the collective interest of all the agent's selves over time and that benefits must be fairly distributed among selves. McClennen & Shapiro, *supra* note 36, at 367. It might follow that if the agent, as now constituted, is viewed as embodying all her past and present selves, she could will action on behalf of the group. Will, however, is most plausibly understood as the will of the agent's current self alone.

⁵⁵ In fact, Bratman makes an argument very similar to this to show that his approach to practical rationality does not suffer from the classic dilemma of rule-utilitarianism. “[Rule-utilitarianism] sanctions utilitarian reasoning concerning rules but does not concerning particular acts. But given its commitment to the former it may seem unclear how it can block such reasoning in the latter case, the case in which we are assessing particular acts.” Bratman sees no similar problem in his approach to practical rationality “because this is only an account of the rationality of an agent for (non)reconsideration [of intentions] that is *not* based on present deliberation. . . . In the sort of case the present account [addresses] there is no need to block direct consequential reasoning by the agent concerning his particular case of (non)reconsideration; for in the case in question there is no deliberation at all about whether to reconsider.” BRATMAN, *supra* note 34, at 209–10, 212.

is simply to follow the rule based on her prior intention and her specific intention to act on it.

The line of argument just described, that temporally extended practical rationality bypasses epistemic difficulties, poses the question whether standards of epistemic rationality require the agent to consider current evidence about reasons for action? If epistemic rationality imposes a responsibility on agents to assess available evidence and form beliefs in response, then the agent is epistemically required to reconsider her intention when her evidence suggests that she currently has contrary reasons for action—even if long-term rationality might favor staying the course. If there is no such epistemic responsibility, then the agent can act unreflectively on her prior intentions without offense to epistemic rationality by simply declining to process her evidence.

A strong version of epistemic responsibility would require agents to advert to all evidence in their possession, draw reasonable inferences from that evidence, and form beliefs that fit the evidence.⁵⁶ This standard, however, is overly demanding. Even when evidence is limited to concurrent mental states, we acquire more evidence than we reasonably can process and translate into beliefs. It follows that agents should not be required to advert to all evidence, or at least that not all evidence should require agents to engage in extended inferential reasoning of the sort required to form beliefs about reasons for action. Suppose, for example, that after P has promised to mow Bob's lawn, her daughter asks for help with child care. In this case, it is not particularly onerous for P to reason from her child's request to a belief about current reasons for action. Suppose instead that agent A, standing outside a building, sees a number of people leave the building with dogs. This observation might support a variety of propositions: for example, that the building owner permits dogs on the premises, that a dog show is happening inside, or that the next person to leave is statistically likely to have a dog. A may have no interest in any of these propositions, and so should not have a categorical duty to treat her observations as evidence and form corresponding be-

⁵⁶ Evidence in the agent's possession is commonly understood to include mental states such as perceptions, conscious memories, conscious beliefs, and background beliefs that make sense of this evidence. See, e.g., Conee & Feldman, *supra* note 43, at 219–41 (defending a strict definition of evidence by a process of elimination). Others take a broader, "direct realist" view of evidence that includes external facts available to the agent. See, e.g., Thomas Kelly, *Evidence: Fundamental Concepts and the Phenomenal Conception*, 3 PHIL. COMPASS 933, 950 (2008). Timothy Williamson defines the agent's evidence as what the agent knows, thus importing a requirement of truth. TIMOTHY WILLIAMSON, KNOWLEDGE AND ITS LIMITS (2000) (taking a strict view).

liefs. In most circumstances, she should be able to tune the evidence out. Accordingly, I will set this version of epistemic responsibility aside.⁵⁷

One alternative is to reject the possibility of epistemic responsibility to respond to evidence. Epistemic rationality, on this view, requires only a reasonable fit between evidence the agent actually considers and beliefs she elects to form in response. Any additional requirements pertaining to gathering and processing evidence or forming beliefs are practical or moral requirements rather than epistemic requirements.⁵⁸ An approach that allows agents to ignore any and all inconvenient evidence, however, seems to miss an important component of rationality in belief.

A compromise position would impose a limited responsibility on agents to advert to and process evidence when the evidence tends to defeat a belief the agent currently holds. According to this formulation of epistemic responsibility, agents do not have a general epistemic responsibility to advert to evidence, draw inferences from evidence, or form beliefs. They must, however, process evidence in their possession that challenges their existing beliefs.

Richard Feldman endorses a standard of epistemic responsibility much like this.⁵⁹ Feldman states initially that epistemic rationality means only that whatever beliefs the agent forms must fit the agent's evidence, but then adds that agents may also be required to process evidence that threatens to defeat their current beliefs.⁶⁰ In Feldman's example an agent believes on limited evidence that ginkgo biloba will improve his memory.⁶¹ The agent then notices the following article title on the cover of a

⁵⁷ This example appears in an article by Adam Leite. See Adam Leite, *Epistemic Instrumentalism and Reasons for Belief: A Reply to Tom Kelly's Epistemic Rationality as Instrumental Rationality: A Critique*, in 75 PHIL. & PHENOMENOLOGICAL RESEARCH 456, 457 (2007).

⁵⁸ See Conee & Feldman, *supra* note 43, at 178–79. Feldman states that “. . . if S has any doxastic attitude at all toward p at t and S's evidence at t supports p, then S epistemically ought to have the attitude toward p supported by S's evidence at t.” The antecedent “if S has any doxastic attitude . . . toward p” is intended to establish that S is under no obligation to form a doxastic attitude. See *id.*

⁵⁹ See *id.*

⁶⁰ *Id.* at 188.

⁶¹ *Id.* at 186–88. Similarly, Alvin Goldman, who normally equates epistemic rationality with the reliability of the cognitive processes by which the agent formed her belief, acknowledges that otherwise rational beliefs are defeated if the agent's current evidence indicates that they are wrong. See Alvin Goldman, *Toward a Synthesis of Reliabilism and Evidentialism*, in EVIDENTIALISM AND ITS DISCONTENTS 254, 275–76 (Trent Doherty ed., 2011).

Goldman's example is a man who reads the morning weather forecast and rationally forms a belief that the weather will be sunny in the afternoon. That afternoon he is caught in a rainstorm but declines to engage in any process of belief revision, ignores his current experiential evidence, and continues to believe that it is sunny. Goldman says that although the agent's belief initially was justified under reliabilist standards, it is now defeated by the agent's current evidence and can no longer be justified. This suggests not only that there are evidential constraints on reliabilism but also that agents have some epistemic responsibility to advert to evidence of which they are aware, at least when the evidence appears to defeat an existing belief. *Id.*

credible magazine: “Ginkgo Shown to Be Ineffective.”⁶² At this point, the agent has evidence contrary to his belief and ought to give up the belief. But an article entitled “Some New Information on Ginkgo” would not have the same effect.⁶³

The motivation for this standard is epistemic rather than practical. The standard comes into play when the agent’s evidence suggests that her current set of beliefs contains errors. Its effect is to restore epistemic order by correcting the erroneous beliefs. A defeat-based standard is also fairly modest, in comparison to the standard of strong epistemic responsibility described above, because it is triggered by existing beliefs.⁶⁴

At first glance, epistemic responsibility to advert to evidence that defeats existing beliefs seems well-suited to the context of promising. At T_1 , when P promises Bob that she will mow his lawn at T_2 , she believes at least three propositions. First, she believes that, given the spectrum of possible circumstances between T_1 and T_2 , her best course of action is to make a binding promise. Second, she believes that there may be circumstances in which the right thing to do at T_2 , all things considered, will be to breach. Third, she believes that she may not correctly identify the cases in which she should breach: her case-by-case judgment is unreliable and will systematically favor a breach. Based on these beliefs, she forms an intention to treat her promise as an exclusionary reason for action. In doing so, she is practically rational in the temporally extended sense.

At T_2 , P has new evidence indicating that she should break her promise and spend Saturday helping her daughter with child care. Her practically rational intention to treat her promise as an exclusionary reason for action dictates that she should ignore this evidence and mow and her reasonable dispositions favor honoring prior intentions of this type. The standard of epistemic responsibility just described, however, suggests that she should attend to her evidence and revise her intention. If instead she proceeds to act unreflectively on her practically rational intention, she will find herself in a state of epistemic irrationality.

In fact, this conclusion may be too quick because P’s new evidence does not actually contradict the beliefs she formed at T_1 and relied on to support an exclusionary promise. P believed at the outset that there would be circumstances in which the right thing to do, all things considered, would be to breach. She believed also that, given the range of pos-

⁶² Connee & Feldman, *supra* note 43, at 187.

⁶³ *Id.*

⁶⁴ Feldman makes the point that challenge to prior beliefs should be fairly obvious. *Id.* at 186–88. In most cases involving promises, this criterion will be met whenever the agent confronts a situation that brings her prior intention into play in unexpected particular circumstances. Connee & Feldman, *supra* note 43, at 186–88.

sible circumstances between T_1 and T_2 and the unreliability of her own case-by-case judgment, the best choice at T_1 was to make and later act on a binding, exclusionary promise. Her new evidence is consistent with each of these beliefs. Technically, therefore, the standard just described, placing epistemic responsibility on agents to process evidence that defeats current beliefs, does not apply. Accordingly, it may still be open to P to ignore her new evidence and complete the long-term practically rational course of action that ends in her performing her promise.

The defeat-based standard of epistemic responsibility under consideration, however, can easily be reformulated to include an epistemic requirement to consider current reasons to break a promise. The underlying problem in the context of promising is what epistemic responsibilities come into play when an agent moves from generalized beliefs to more particular beliefs, acquiring new clarifying evidence about the circumstances to which her prior general beliefs apply. The beliefs that led P to make a binding promise may still hold true as a general matter, and yet need revision to account for previously unknown particulars. Another way to characterize the problem is that although P's new evidence does not contradict her current belief that she ought generally to treat her promise as an exclusionary reason for action, it suggests a better belief: she ought generally to treat her promise as exclusionary, but not in this case.

Modified to reflect this type of epistemic error, the defeat-based standard of epistemic responsibility would require the agent to respond to evidence showing that her existing general beliefs need revision to accommodate her present circumstances. From an epistemic point of view, treating her promise as an exclusionary reason for action is generally the right choice, but in the circumstances she now faces it is not. A standard requiring belief-revision in cases of this type serves the same epistemic values as the more typical defeat-based standard of responsibility first described, which requires agents to process evidence that may defeat existing beliefs. In both cases, responding to evidence and modifying prior beliefs corrects a problem that has come to light and restores the integrity of the agent's set of beliefs.

Of course, acting on the modified belief may not be practically rational over the long run because the agent may be wrong. Her evidence may be incomplete and her inferences may be imperfect. This is why she formed a general intention to treat her promise as exclusionary in the first place. Epistemically, however, what counts is that the agent's existing beliefs should be revised to match her reasonable assessment of current evidence that bears on their justification.

Although this revised version of a defeat-based standard of responsibility has practical repercussions, from an epistemic standpoint, its ef-

fects are limited. Agents are not required to advert to all evidence in their possession, or to draw all possible inferences from evidence, or to form all beliefs that match their evidence. They are not, for example, required to think through the implications of a random observation about people and dogs. As in the case of defeating evidence that Feldman describes, the agent is required only to draw ordinary inferences from readily accessible evidence, in order to determine how a prior belief holds up in a new case. Thus, the principal drawback of this approach is not the burden imposed on agents but the inconvenient consequences for temporally extended practical rationality.

Thus, under plausible assumptions about epistemic responsibility, epistemic rationality appears to conflict with practical rationality in the context of promising. Long-term practical rationality favors unreflective performance of promises, while epistemic rationality requires reflection on new evidence and, if indicated, revision of the promisor's intention to perform. A promisor who treats her promise as an exclusionary reason for action may be practically rational, but she is not epistemically rational; thus, she is not fully rational.

CONCLUSION

Exclusionary reasons for action provide a sensible explanation for the bindingness of promises, if it is rational for agents both to intend their promises to operate as exclusionary reasons and to treat them as exclusionary when the time comes to perform. There are a number of practical reasons to intend one's promises to have an exclusionary effect. The more difficult question is whether an epistemically rational agent can maintain and act on such an intention.

Theories of temporally extended practical rationality such as Bratman's provide support for the practical rationality of treating promises as exclusionary reasons for performance.⁶⁵ They do not, however, offer a full solution the problem of rationality because, on a plausible understanding of epistemic rationality, agents cannot treat promises as exclusionary without engaging in epistemic irrationality. At best, promisors who perform their promises without reflecting on current reasons for action are not fully rational. If epistemic rationality is foundational to practical rationality, promisors who perform without reflection may also be irrational overall. Thus, exclusionary promises are a good idea that does not quite work.

⁶⁵ See BRATMAN, *supra* note 34.

