MENACHEM FISCH

and

YITZHAK BENBAJI

THE VIEW

FROM WITHIN

Normativity and the Limits of Self-Criticism

University of Notre Dame Press

Notre Dame, Indiana

© 2011 University of Notre Dame
CHAPTER ONE

Setting the Stage

The Problem with Rationality

Two Concepts of Rational Action

Theories of rational action fall under two main headings. The first pro-
nounces an action rational if, given the circumstances, it is perceived to
be the right or proper thing to do. To act rationally is to act in ways that,
given the actor’s needs, desires, and, situation, meet expert approval. On
this view, what is scrutinized in assessments of rationality are the merits
of the actor’s actual performance rather than the quality of the data at
her disposal, her level of expertise, or the quality of her reasoning. It is the
end result that counts. To be considered rational in this sense, it is enough
for an act to be deemed effective or worthwhile. What is appraised is per-
formance, not agency. Action taken or withheld thoughtlessly, instinctively,
or unwittingly may also be considered rational as long as it is deemed ap-
propriate. Rationality thus construed is, therefore, not limited, in prin-
ciple, to human action. Instinctive animal behavior can be deemed ra-
tional in this respect. Likewise, properly functioning computer programs,
automatic pilots and thermostats, programmed to “sense,” “assess,” and re-
spond to a situation effectively, are often said to perform rationally with-
out doing too much violence to the term. It is a notion of rationality used
widely in talk of skillful feats one trains to perform instinctively without
having to ponder on them each time anew. It matters little how the skill was originally acquired. Often, as Michael Polanyi famously argued, the learning process may itself remain tacit.\(^1\) Most of us learn to perform highly complex feats of logical reasoning, for example, without ever having studied logic as such. The move from tacit human knowledge acquisition to preprogrammed, hardwired animal and artificial skilled response is then quite natural.

To paraphrase Bernard Williams’s well-known distinction, it is not \(X\)’s reasons for \(\phi\)-ing that are being judged but the extent to which \(\phi\)-ing is considered the reasonable thing for \(X\) to do.\(^2\) Rationality on such a showing is hence a matter of situation analysis. It has to do with the weighing of possibilities and calculation of possible outcomes. Due to the supposedly calculable, algorithmic nature of its underlying logic, this notion of rationality has attracted enormous attention in recent years, especially among decision and game theorists.

Thus construed, rationality still requires reflection and sound assessment, but it is not the actor who is required, or, in most cases, even qualified to do so. To be considered rational she is expected to act in conformity with what is considered appropriate by the best means available but without necessarily having to have worked it out for herself. Indeed, unless the actor happens to be the referring expert herself, her own account of her actions will be of little if any relevance to their rationality.\(^3\)

Nor need the required approval be unanimous, or granted immediately. Approval, and with it assessments of rationality, may be conceded by some and not by others and, even then, only long after the event. Rationality, in this sense of the term, is time- and context-dependent, in principle, and is hence interestingly relativized despite its supposedly objective, calculable nature. Actions deemed inappropriate and hence irrational by some may come in time to be considered fitting and perfectly rational by others—and vice versa.

Although the rationality of the type of action in question may have been ascertained long before the specific act under consideration was performed (as in the case of standard countermoves in chess, or an appropriately applied well-honed mathematical technique), evaluations of this first category of rationality are always backward-looking. What is deemed rational is forever what the actor ended up doing or refraining from doing.
One does not embark on a process of deliberation that can be deemed rational before it is completed and its outcome independently assessed. Rationality, in this sense of the term, is always decided after the event. It is not a prospective evaluative category of acting but a retrospective evaluative category of types of actions. What makes for the rational is not what goes into practical reasoning, if you wish, but what comes out. It is the move the actor actually makes that is assessed regardless of the manner in which she reasoned or conducted herself prior to making it.

The second approach to rationality is less liberal. It insists, as Robert Brandom puts it, on distinguishing human rationality from “the mere reliable differential responsiveness that we share with non-conceptual creatures such as pigeons—or as far as that goes, with photocells and thermostats.” It demands of moves deemed rational to be more than merely right or efficient. It is trickier than the first, and is less prone to calculation. To be considered rational, it is not enough for a move to gain retroactive approval; an action is required to be the outcome of considered deliberation. Rationality is made to turn not on what the actor ends up doing but on the manner in which she arrives at her decision to do so—gauging the quality of reasoning culminating in the act rather than the merit or effectiveness of the act once made. Here, by contrast, agency reigns supreme. Thoughtless action, appropriate, effective, and meritorious as it may turn out to be, will never be deemed rational. There are other evaluative categories by which to commend action that merely fits the bill. Rationality is here construed as an evaluative category reserved exclusively for considered action, action taken self-consciously and reflectively.

Rationality’s Authority

But by what standards are we to judge an actor’s considered deliberation? The answer, obviously, is by appealing again to expert opinion. An action will be deemed rational only when taken by virtue of such feats of practical reasoning that meet authoritative approval. The canons of rationality appealed to here differ substantially from those of the first-order situational analyses that gauge the first model. Here the actor is not expected to act merely in conformity with the outcome of expert diagnosis, but to emulate expert diagnosing. To paraphrase Bernard Williams
again, to be deemed rational, it is not enough for a move to be considered reasonable by approved standards; it is required to be reasoned for approvingly. The first sense of rational conduct required actors merely to act as would an expert, the second requires of them to reason as experts do.

The two notions of rationality differ significantly. At best, the class of activity deemed rational by the second forms but a tiny subset of that deemed rational by the first. (But even that is not entirely true, for according to the second, an action may be considered rational even when unsuccessful, or deemed mistaken, which for the first would be unthinkable.) The demands on the actor are in a sense far more stringent in the second than in the first. But the two do have one important feature in common: they are both ultimately gauged by appeal to authoritative approval. The foci of approval are very different in each case, and, consequently, so are the criteria by which it is leveled. Still, gauging the rationality of a move by appeal to accepted standards renders rationality inherently context-dependent. It matters little if the standards invoked are the accepted canons of appropriate action or those of appropriate reasoning. In either case, rationality becomes a matter of conforming to the currently accepted standards and hence, by definition, wholly dependent on time, place, and opinion.

Indeed, why not? Do not all our evaluative categories invoke standards and norms, and are not all our standards and norms culturally determined? Is it not obvious that action considered appropriate in one context may well be deemed unfounded in another; that conduct deemed reasonable in one cultural setting may be considered outrageous in another? Needless to say, the arguments for the relativity of the normative in this regard are powerful. With respect to rationality, however, there is a price to pay.

Second-Order Pondering

To see this we need to take a closer look. As noted, the context-dependency of the first notion of rationality makes perfect sense. If the rationality of a move is thought to be calculable by the most recent decision-theoretic or game-theoretic approach, it should be expected to change as the theory is further developed or replaced. If rationality amounts to acting in accord
with expert opinion, then, by definition, it is a matter of opinion. What allegedly saves rationality from becoming merely a matter of opinion is, of course, the fact that it is made to turn, not just on anyone’s opinion, but on that of reliable assessors who are expected to form their opinions rationally. But if rational action is action taken in conformity with approved standards of performance, how is the work done on those very standards to be rated? Since standards of adequate performance of both acting and reasoning undergo changes from time to time, by what means, and with respect to which measures, can their second-order scrutinization and replacement be deemed rational, arbitrary, or capricious? This is where the first notion of rationality converges on the second, and where both become problematic.

Our reason for focusing on expert, rather than personal, common, communal, or majority, opinion is to lay special emphasis, if only metaphorically, on the special role played by informed specialists in devising, reviewing, and occasionally revising the standards, norms, and methods of their fields of expertise. Even the most avid supporters of the first notion of rationality will allow thoughtless, mindless action to be considered rational only if it can be said to merit the authoritative approval of someone who has competently thought it through. Unconsidered action, or activity performed thoughtlessly by an entire community that is merely following custom, instinct, or habit, will not be considered rational by either model. The two approaches differ with respect to the level of reasoning, consciousness, and agency they demand of whoever (or whatever) performs the action in question. But both agree that to be deemed rational, an act must be actively and consciously approved. Rationality, on either showing, would seem to require more than knowingly or unwittingly following a rule or a ruling; it requires that the rule or ruling followed be fittingly approved and endorsed.

An approved norm, rule, or ruling is a considered one, one that at some point was weighed and deliberated, compared and tested for situations and tasks resembling those at hand. What makes for rational action, according to both understandings of the term, is the fact that it is reasoned. Although we side with the second approach—insisting that for action to be deemed rational the actor herself be required to perform the reasoning—we shall press the matter here no further. Rather than
attempt to adjudicate between theories of rationality, let alone propose one of our own, this book, as stated at the outset, focuses on what we hold both approaches to rationality have in common: namely, a requirement that the rational turn centrally on some form of second-order, deliberative assessment and what Christine Korsgaard terms “reflective endorsement” of the sort exemplified most fully in prudent expert evaluation.\(^8\)

The phrase “have in common” could be misleading, though. For we believe that this type of reflective deliberation is more than something theories of rationality just happen to share. On the contrary, it is our contention that the various approaches to rationality exhibit wide if at times implicit agreement that this type of reasoning is, in the last analysis, rationality’s widely shared defining feature. It is where humankind’s capacity to reason (minimally) enters the picture to make the rational what it is and to mark it off from what it is not.

The sort of evaluative approval to which we refer requires a stepping back from the well-honed ways of old with a view to considering them to some degree anew. If rational action is action taken with a view to effectively meeting a need or accomplishing a goal, then to approve such action amounts, minimally, to critically reviewing the means selected to do so. A more robust evaluation might question the needs and goals themselves.\(^9\) But let us concentrate for the time being on the minimal case.

This type of reasoning, located at the heart of rational endeavor, aims at carefully reviewing the actor’s circumstances, values, assets, and goals, with a view to assessing the aptness of her response to the specific task at hand.\(^10\) If the exercise is not an empty one, it will always involve, if but fleetingly, a critical weighing of alternatives. In routine situations it will amount to no more than a nodding acknowledgment of the selection by the actor of the standard tool for the job. But not all situations are routine. Some of the time much more will be at stake and much more will be involved. Each time the actor’s specific circumstances, needs, or purposes defies standard procedure, standard procedure will itself become the object of reflection. Most often even then little more will be required than adjusting and tuning standard tools and methods to meet specific conditions in routine and well-rehearsed ways. But often it is not. And when it is not, that is when the standards and the norms constitutive of the field are liable to be called into question. These are moments
when experts are required to reflect critically on the tools of their trade rather than expertly apply them.

In facing such outstanding problems they will often have to rely on the findings of sister disciplines. Peter Galison has drawn attention to the way well-managed “trading zones” are established wherein practitioners of different disciplines communicate and “trade” by means of suitably watered-down, pidgin versions of their different professional vocabularies. But dependent as a field may be on other fields for solving problems, it is seldom the case that the problems themselves are comprehended outside it. No one is better placed than the foremost practitioners of a field to encounter, experience, and acknowledge difficulties that arise within it, just as it is they who will subsequently be charged with the task of considering, debating, and eventually approving their solutions.

Needless to say, fields operate by means of norms, standards, and procedures some of which are more easily modified than others. Some fields of activity—those that constitute language games in the more literal sense of the word, such as in sports and some arts—maintain rigid rules and regulations in order to increase the challenge and the competition artificially. But the norms, standards, and procedures that regulate the activity in most fields are not maintained merely for the sake of making the play more difficult or challenging, and yet they are not regarded by practitioners as sacred absolutes imposed from without. Rather, they are endorsed and maintained normatively as well-considered means to well-considered ends. And just as they are seriously endorsed and maintained as such, they, too, are liable to come under serious review from time to time in response to problems encountered.

It should not come as a surprise that the mode of reasoning that makes for rationality thus turns out to be the very same mode of reasoning that accounts for the studied operation and the reasoned fashioning and refashioning that occurs in all fields of considered human endeavor. If to act purposefully is to apply a means to an end, and if what makes for the rationality of such a move is that it be competently reviewed and approved, then the two modes of reasoning are not merely similar, but in this sense fully coincide. Competent assessors of rational action are experts, by definition. And the role they are expected to play in assessing
such action is precisely the same role they are expected to play within their respective fields. The point we wish to emphasize, however, is less the actual agreement between the two roles, which is pretty obvious, than the wide spectrum of activity of which they consist. This, we have seen, extends from the nodding approval of standard uses of standard tools and methods, at its lowest point, all the way up to the studied, at times anxious, rethinking of the field’s very foundations in the face of outstanding difficulties.

Note, however, that we do not demand that to be deemed rational action must be approved or reasoned for correctly in any absolutist or foundationalist sense of the term—only that it be approved, reasoned, questioned, and reflected on authoritatively, that it be normatively endorsed. Rationality thus construed would hence seem quite susceptible to the cultural relativity of the normative central to much latter-day philosophy of language, mind, self, and culture. Or so it would seem. We shall return shortly to the three master ideas that animate this study: (a) the central identification of an action or a belief’s rationality, with it being considered, that is, embarked on or endorsed for a reason; (b) the essential normativity of such reasoning; and (c) the inevitable role of normative self-criticism in the rational modification or replacement of normative commitment. But first a word on relativism, or what we prefer to refer to as Normative Diversity.

Normative Diversity: An Introduction

In what follows, we take relativism in the sense endorsed and held in esteem by Richard Rorty, as opposed to the sense he rejects (despite his objection to the very use of the term). The notion of relativism in which we are interested, and with which we sympathize, differs from the thesis sometimes termed “cultural” or “moral relativism,” alleged to claim that any viewpoint is as good as any other. Such a position begs the question, if only by presupposing a transcultural notion of ‘good’. According to the more stringent position we shall be interested in, such a “view from nowhere”—to use Thomas Nagel’s oxymoronic metaphor from which the title of the present study takes its cue—is humanly unavailable. A
culture, a language game, a form of life, a normative outlook, it argues, can only be viewed from within, or, alternatively, from the necessarily biased perspective of another. From these two humanly available perspectives (which roughly map onto Korsgaard’s distinction between first- and third-person perspectives with regard to the normative), mutual understanding and normative evaluations and preferences are quite coherently forthcoming but, by definition, are never “objective” in the transcultural, framework-independent sense of the term. Relativism, as we refer to it here, is the position claimed to follow from the very idea by which cultural relativism is rejected, namely, the denial of the availability of a normative vocabulary that “has nothing to do with agency, values or interests,” and that can, therefore, not be shared by radically different normative outlooks. It is the position we dub in what follows “Comparative Normative Irrealism,” one that denies thin normative concepts like ‘true’, ‘good’, ‘despicable’, ‘rational’, and ‘sound’ the capacity to compare and rank diverse normative systems, in a manner acceptable to both. “Dewey,” writes Rorty,

thought that the Kantian notion of “unconditional obligation” . . . could not survive Darwin . . . For him, all obligations were situational and conditional. This refusal to be unconditional led Dewey to be charged with ‘relativism.’ If ‘relativism’ just means failure to find a use for the notion of ‘context-independent validity,’ then this charge was entirely justified.

Such a position, it is argued further, entails that meaningful discussion of and reflection on one’s own normative vocabulary (or language game, “system of belief,” “form of life,” or “final vocabulary” as such frameworks are variously termed) is necessarily determined by and contained within it to the extent that, as Rorty rather forcefully puts it, “all we can allow ourselves” is to “limit the opposition between rational and irrational forms of persuasion to the interior of a language game, rather than try to apply it to interesting and important shifts in linguistic behavior.” It is only “within a language game, within a set of agreements about what is possible and important, [that] we can usefully distinguish reasons for beliefs from causes for belief which are not reasons.” Relativism, as we take it here, is the view according to which, to paraphrase Hilary Putnam,
“elements of what we call ‘language’ or ‘mind’ penetrate so deeply into what we call ‘reality’ that the very project of representing ourselves as capable of rationally reviewing and revising those elements “is fatally compromised from the start.”23

A Two-Step Argument

In other words, such relativist construals of the normative, it is argued, severely constrain the extent of rational review and revision to which a normative framework can admit from within. It is a two-step argument. The first consists of the sobering neo-Kantian realization that all meaningful normative discourse is determined by, and hence relative to, the particular normative vocabulary it draws on. Following such thinkers as Rorty,24 Davidson, Williams, Putnam, Michael Friedman, Charles Taylor, Michael Walzer, and Brandom, whose work we shall discuss in some detail, we fully concede this first step. In doing so, we shall be arguing against the opposition, leveled most interestingly from more orthodox latter-day Kantian quarters, such as in the work of Korsgaard and John Rawls and his school and, to a more limited extent, John McDowell’s recent work.

The second, more radical step of the argument urges us to realize that by virtue of their constitutive role in the discourses they support, our normative vocabularies are themselves rendered largely immune to normative critical appraisal from within those discourses. In being constitutive of all rational reasoning and reckoning within their scope, it is argued, they defy becoming the objects of rational reasoning and reckoning from within. Therefore, though obviously the products of human endeavor, they cannot be considered the products of rational human endeavor. Scrutinized by their users, a normative vocabulary may be found to harbor inconsistencies, incoherence or, from a Taylorian or Brandomian perspective, felt to be in need of further explication. Problems may also arise regarding the relative priority ascribed to norms. But by virtue of its constitutive role in determining the good, the right, and the appropriate, a normative vocabulary, it is argued, cannot be deemed from within to be bad, wrong, or unfitting. A final vocabulary—to use Rorty’s coinage—namely, the set of words with which people justify their actions, and about which “if doubt is cast . . . their user has no noncircular
argumentative resource,”25 is not and, in principle, cannot be the outcome of considered, reflective deliberation of the kind we locate at the heart of rationality.

Many oppose the type of relativism of which we speak—some dismissively, some thoughtfully. But many concede it—more, perhaps, than one would have thought, as we shall argue in some detail in upcoming chapters. Among those who concede it, none, to the best of our knowledge, have taken issue with the second step of the argument, although only very few acknowledge it explicitly; Rorty and Davidson stand out in this respect.26 Work on rationality among those we class as normative irrealists shows little interest in and exhibits virtually no discomfort with the idea that our normative vocabularies elude rational scrutiny. The exceptions, as we shall see, are Friedman’s work on the rationality of scientific framework transitions and Brandom’s common law, Hegelian model of “historical” rationality.27 However, even their notions of rationality, as we shall argue in some detail, are wholly lacking of the normative self-critical aspect of rational evaluation on which we insist.

Put briefly, the aim of the present study is to explore the possibility of resisting the second step of the argument while fully endorsing the first; to salvage, in other words, a viable notion of normative self-criticism, applicable to both intersubjective and intrasubjective normative discourse, yet without resorting to an external or comparative normative foothold. Normative vocabularies, we shall show, are susceptible to normative criticism from within, despite their essential incomparability to sufficiently diverse alternatives. But before explaining in more detail why, and outlining how we intend to do so, a little more needs to be said about step 2 in order to get a better sense of the difficulties involved in contesting it. We do so by way of an introduction by appealing to the work of its most outspoken and eloquent advocate, Richard Rorty.

Rorty’s Relativism

Richard Rorty, a primary proponent for the two-step argument, does not regard its conclusion in any way problematic. On the contrary, the “ironist” strong poet, whose image he portrays and promotes in _Contingency_,
Irony, and Solidarity, and with whom he fully identifies, is someone who knowingly conducts herself in accord with and in full awareness of this constraint on her rationality. She is someone who fulfills three conditions:

1. She has radical and continuing doubts about the final vocabulary she currently uses . . . ;
2. She realizes that arguments phrased in her present vocabulary can neither underwrite nor dissolve these doubts;
3. Insofar as she philosophizes about her situation, she does not think that her vocabulary is closer to reality than others, that it is in touch with a power not herself.

She is portrayed as self-critical but only in the sense of tracing the outer borders of self-criticism. Aware of her final vocabulary’s inherent contingency, she doubts it persistently and spends “her time worrying about the possibility that she has been initiated into the wrong tribe, taught to play the wrong language game.” However, and this is the crucial point, “she cannot give a criterion of wrongness” and thinks, therefore, that there “is no reason to think that Socratic inquiry into the essence of justice or science or irrationality will take one much beyond the language games of one’s time” (74–75). She therefore experiments in fashioning new vocabularies by blindly casting around for new metaphors to redescribe herself and her world. Unable to question, she dabbles; unable to troubleshoot, she tinkers. She gropes blindly because she assumes, with Rorty, that from the vantage point of her current vocabulary—which is the only committed vantage point available to her—it is impossible to articulate what might be wrong with it, or in what ways it might be improved, and that, subsequently, “the creation of . . . a new vocabulary, will have its utility explained only retrospectively” (55), and, again, only from within. Rorty’s ironist aims ultimately to improve on her vocabulary, rather than merely change it for another, but can only do so aimlessly. She can envisage change, but, unable to set herself a prospective goal, other than change itself, she cannot envisage progress. All she can do is to do her best to be original, to keep telling her story differently. She works on her vocabulary but cannot appraise it critically. She can, therefore, not proceed rationally in improving it, in the sense we have described and shall explain further below. She can only blindly redescribe with the hope of hitting unwittingly on an option that in retrospect she will find worthwhile.
Rorty refers frequently and glowingly to the work of Thomas Kuhn, whom he describes as “one of [his] idols,” “a great philosopher,” “the most influential . . . to write in English since the Second World War.”

Most of the time Kuhn is praised for redescribing science to show that “the subject of truth claims cannot be a relation between beliefs and a putatively mind-independent or ‘external’ world,” thus contributing to the idea central to Rorty’s entire project that “correspondence to reality is a term without content.” When he does refer to Kuhn’s description of paradigm shifts he is less clear than one would expect, however, tending to slide too easily into Kuhn’s collectivist idiom and shift attention from the “strong poets” responsible for fashioning the paradigm shift to the community that ends up endorsing it. This is to blur the difference between rational inquiry and rational acceptance; between perceiving and responding to problems, on the one hand, and being willing to entertain their solution once proposed. Rorty thus creates the false impression that the issue is essentially sociological. “As Kuhn argues in The Copernican Revolution,” Rorty writes, “we did not decide on the basis of some telescopic observations, or on the basis of anything else, that the earth was not the center of the universe, that macroscopic behavior could be explained on the basis of microstructural motion, and that prediction and control should be the principal aim of scientific theorizing.” One wonders who the “we” are to whom Rorty refers. He quickly clarifies: “Rather, after a hundred years of inconclusive muddle, the Europeans found themselves speaking in a way which took these interlocked theses for granted. Cultural change of this magnitude does not result from applying criteria. . . . We should not look within ourselves for criteria of decision in such matters any more than we should look to the world.”

The question addressed here is how to explain how a community, or a society at large (“the Europeans”), acquires and internalizes a newly fashioned final vocabulary after it was formed rather than the very different and far more radical question, so central to his book, regarding the manner in which such radically new vocabularies are fashioned by their authors in the first place. Rorty’s reference here to “the Europeans,” rather than to the individuals who set the scientific revolutions he mentions in motion, creates the false impression that his reference to Kuhn pertains to the first rather than to the second question. The passage immediately preceding the one quoted, goes beyond Kuhn to include
paradigm shifts outside the sciences but remains similarly misleading: “Europe did not decide to accept the idiom of Romantic poetry, or of socialist politics, or of Galilean mechanics. That sort of shift was no more an act of will than it was a result of argument. Rather, Europe gradually lost the habit of using certain words and gradually acquired the habit of using others.” Discerning how professional communities and the educated public at large lose and acquire such habits of thought and speech is a complex and fascinating story in itself, of course, to which we return if briefly in chapter 9. But the real challenge to rationality is that of Galileo, the poets, and the social reformers themselves, who spent a lifetime deliberating, contriving, and perfecting these new vocabularies. Can Galileo’s creative enterprise be deemed “no more an act of will than it was a result of argument”? Can his lifework also be described as a matter of merely losing “the habit of using certain words” and acquiring “the habit of using others”? When these early passages of Contingency, Irony, and Solidarity are reread in the light of the self-portrayal of Rorty’s ironist developed later in the book, the answer turns out to be a clear and disturbing “Yes!” And indeed, a little later in the first chapter, he says so explicitly. Wittgenstein’s analogy between a vocabulary and a tool, he says, “has one obvious drawback.”

The craftsman typically knows what job he needs to do before picking or inventing tools with which to do it. By contrast, someone like Galileo, Yeats or Hegel . . . is typically unable to make clear exactly what it is that he wants to do before developing the language in which he succeeds in doing it. His new vocabulary makes possible for the first time a formulation of its own purpose.

Only after a new vocabulary is up and running, as it were, can its creators appreciate retrospectively what it is good for and how it fares better than the vocabulary it replaced. Although few would disagree that there is always more to a tool than what goes into it and that it is never possible to anticipate all its uses before it is actually employed, Rorty’s position is far more radical. Only after an alternative vocabulary is endorsed and employed can the problems solved by its endorsement even be formulated. A functioning final vocabulary can only be deemed normatively wanting from the perspective of another.
In other words, a final vocabulary cannot be normatively troubleshoot and modified from within but only by someone working from the outside, someone whose final vocabulary it no longer is. A person can be “ironically” suspicious of her final vocabulary and desire to experiment with other ways of life, as Rorty’s ironist is. But she can never do so for a reason. One is unable ever to state what is, or might be, wrong with one’s normative commitments (other than to find them incongruous with other of one’s norms or in need of further clarification). In short, a language game cannot be reflected upon, it cannot admit of normative shortcomings, cannot be criticized, and cannot be improved upon from within, except in the trivial sense of troubleshooting for consistency, coherence, or clarity. Language games may be fashioned by humans but never for a purpose other than for their sheer novelty. And thus, as Rorty is the first to admit, precious as it may be, our admirable capacity for self-criticism and self-improvement is required by his philosophy to stop short of its seemingly profoundest, to be replaced by the aimless experimentation of clueless irony.

At one point Rorty briefly explains why. Quoting Davidson’s “Paradoxes of Irrationality,” and alluding to the distinctions between thick and thin normative concepts as well as to that, late of Frankfurt, between first- and second-order desires, Rorty presents a strictly coherentist account of in-house critical reflection. He notes that if we reserve the term rational to mean something like “internal coherence,” . . . then we shall be forced to call “irrational” many things we wish to praise. In particular we shall have to describe as “irrational” what Davidson calls “a form of self-criticism and reform which we hold in high esteem, and that has even been thought to be the very essence of rationality and the source of freedom.”

The process by which “a person forms a positive or negative judgment of some of his own desires,” and acts to change them, he explains, is limited to acts of self-criticism and self-improvement pertaining to a “higher-level” desire capable of mediating and rationalizing the contesting lower first- and second-level desires. However, working up the hierarchy one necessarily reaches a point where
the only candidates for such highest-level desires are abstract and empty as to have no mediating powers: They are typified by ‘I wish to be good’, ‘I wish to be rational’, and ‘I wish to know the truth.’ Because what will count as ‘good’ or ‘rational’ or ‘true’ will be determined by the contest between the first- and second-level desires, wistful top-level protestations of goodwill are impotent to intervene in that contest. (49)

The possibility of rational appraisal doing real normative work swiftly evaporates as normative concepts become “thinner.” And thus, just at the point that our highest and most relished norms and standards “thicken” out to acquire the precise meanings they have for us, they freeze and fossilize, and, for lack of independently meaningful higher standards by which to judge them, are rendered immune to the normative scrutiny of those whose form of life they govern.

Given these constraints on rational self-criticism and self-improvement, Rorty is left no choice but to portray the self-consciously critical pathbreaking work of a Galileo or a Hegel as the outcome of the aimless tinkering of a Rortian ironist. He is aware of how outrageous he must sound but sees no reason to reconsider. He voices the natural objections, only to dismiss them, not by taking anything back, but by rejecting the vocabulary by which they are voiced. Going as it does to the heart of the problem, his comments here are worth quoting at length.

To accept the claim that there is no standpoint outside the particular historical conditioned and temporary vocabulary we are presently using from which to judge this vocabulary is to give up on the idea that there can be reasons for using languages as well as reasons within languages for believing statements. This amounts to giving up the idea that intellectual or political progress is rational, in any sense of “rational” which is neutral between vocabularies. But because it seems pointless to say that all the great moral and intellectual advances of European history—Christianity, Galilean science, the Enlightenment, Romanticism, and so on—were fortunate falls into temporary irrationality, the moral to be drawn is that the rational-irrational distinction is less useful than it once appeared. Once we realize that progress, for the community as for the individual, is a matter of using new words as well as arguing from premises phrased
in old words, we realize that a critical vocabulary which revolves around notions like “rational”, “criteria”, “argument”, and “foundation” and “absolute” is badly suited to describe the relation between the old and the new.42

The problem regarding our critical vocabulary is, of course, not at all that of describing “the relation between the old and the new” but that of describing the effort invested in forming and fashioning the new as an activity thoughtfully and meaningfully performed on, and from within the old. It is not that of overcoming the incommensurability between two normatively incongruous existing frameworks but that of deeming the search or creation of one as rationally motivated while committed to the other.

Nor has the question, as we pose it and shall elaborate further in upcoming chapters, anything to do with “foundations” or “absolutes.” Rational, as we insist on employing the term, is a category of action and agency, not of the relations obtaining between statements, languages, or paradigms.43 The question is not whether the new can be understood or inferred from the old but whether its fashioning can be described as the outcome of the reasoned deliberation of anxious practitioners committed to and working from within the old.

Rorty flatly denies that it can. He is clearly aware that in saying so he is doing violence not only to the term rationality, but to everything we know about the way in which creative individuals critically reflect on, not merely toy with, their norms, their standards, their tools, and their methods. It strikes us not merely as irrelevant or superfluous—or as pointless, as Rorty has it—to describe such “great moral and intellectual advances of European history” as the fortunate falls of aimless, blind tinkering but as downright absurd! Rorty, one feels, should not be allowed to get away with it by merely switching terminologies or refusing to use certain words (a move he makes quite purposefully, one should add, not as a result of blindly toying).44

The question, however, is not a matter of sticking with or abandoning certain words, and certainly not that of approving or disapproving such evasion tactics. The problem is a major one. Rorty’s problem is that, absurd as it may seem to some, the philosophy of language and mind to
which he subscribes will not allow him to view normative vocabularies as subject to reasoned modification or replacement (except, again, in the uninteresting sense of testing for consistency, coherence, or clarity). For him, dropping the troublesome terminology is not an evasion tactic. His philosophy leaves him no choice.

Does our commitment to the rational appraisal of the standards and norms we employ leave us no choice but to drop Rorty’s troublesome philosophy? The answer is neither a simple yes nor a simple no. So far all we have established is that Rorty’s normative relativism is incompatible with the notion of rational appraisal when applied reflexively to the normative vocabulary by which it speaks. If that was all we were claiming, it would have been a matter of deciding to which horn of the dilemma we were more deeply committed. But the matter is not so simple for the simple reason that Rorty’s normative relativism is not an option easily set aside. So far we have stated our own preferences and described Rorty’s contrasting account of the dynamics of normative vocabularies and contrasted it with our own. We have not yet argued for either. It is now time to say a little more about our point of departure, and to examine a little more closely the theories of thought, meaning, and commitment that ground the type of position Rorty represents. Only then will we be in a position to take fuller stock of the dilemma and the prospects of confronting it anew by means of more than a change of words.

The Road Taken

Those who accept (some version of) the first, neo-Kantian step in Rorty’s argument but, like us, are unwilling to accept his radical conclusions have one of two nondissipmissive options. One option is to argue for there being some kind of foothold external to one’s normative vocabulary capable of facilitating its critique from within.45 The modest absolutism premised, for instance, by Dworkin’s Rawlsian critique of Walzer’s interpretivist ethics is one such approach.46 The external foothold, as we shall see, need not be absolutist, however. Michael Friedman’s Dynamics of Reason, though limited exclusively to science, belongs in this subtler category (see chapter 5 below).
The second option is to meet the challenge head-on by opposing Rorty’s conclusions without contesting the premises that support them (or, as described above, by conceding the first step and contesting the second of the two-step argument). This, as we have stated, is the option adopted here.

Our interests naturally lie beyond the specifics of Rorty’s own position. It is the type of challenge his work poses to rationality that animates this study, one we wish first to reformulate in a way that we find most general and effective. Properly constructed, it is a challenge we find deep and far-reaching, to a large extent valid, and not at all easily dismissed. In our reading, it comprises two oft-conflated theses that, we believe, can and need to be properly formulated, analyzed, and distinguished. The first, mentioned briefly above, is the denial of a normative scale on which two maximally coherent yet sufficiently diverse normative outlooks can be ordered in a manner acceptable to both. It claims, in other words, that there is no purely normative relevant comparative property that is shared by significantly conflicting normative outlooks, and, therefore, is capable of ranking them. We call this meta-normative position Comparative Irrealism. Comparative Irrealism, however, entails that because (unbiased) normative comparisons of sufficiently diverse (and sufficiently coherent) outlooks are not forthcoming, an outlook cannot be deemed normatively wanting from within in comparison to a superior alternative. \[47\]

The second claim, to which the present work devotes itself to contesting, is that since an outlook cannot be found normatively wanting by comparison, it cannot be found normatively wanting at all! The logic is simple: if for lack of a suitable comparative normative vocabulary, an outlook cannot be deemed by its adherents normatively inferior to another, then, for the very same reason, neither can they find it normatively wanting in itself. For the question would always be: wanting in comparison to what? A normative outlook may be judged from within to be inconsistent, incoherent, or unacceptably inexplicit, and, to some extent, even lacking in comparison to another. But, lacking a comparative dimension, the argument goes, it cannot be deemed wrong, false, or unjust. And if an outlook cannot be deemed normatively problematic by its own lights, its present normative state can never be deemed to be the outcome or the subject matter of rational scrutiny as defined at the outset.
The ultimate aim of the following pages is to drive a wedge between Comparative Irrealism and the claim claimed to follow from it, according to which all normative criticism is by nature comparative. We thus seek to salvage a robust notion of normative self-criticism while fully conceding the broadly conceived neo-Kantian and neo-pragmatist premises of the first step of Rorty’s argument.

Before outlining the argument unfolded in the following chapters, let us reiterate our point of departure and motivation for taking the road that we take. We set forth from what we characterized above as rational action’s defining feature (even for those who do not require agency in the actual performance of the acts they consider rational), namely, that it is (at least in the last analysis) considered action, action taken on reflection, for which a person is liable to be prompted and to be willing to give his reasons for taking. In other words, we take rational action to be action deemed appropriate. Which, in yet other words, is to take it to be inherently norm governed. From which follows that for an act to be deemed rational, standards of propriety must be in place at least prior to the deeming. It is in this sense that we take normative commitment to be constitutive of rationality. Hence the problem of revising normative commitment rationally.

Moreover, although the point is little stressed in the literature, we insist in addition, as noted above and as the language of the previous paragraph implies, on taking rationality to be, in essence, a prospective category of acting. There is a crucial difference between action opted for for a reason and action taken thoughtlessly and justified only after the event, between the exercise of reason in evaluating a move and in making it. Retrospective justification, or rationalization, can justify an act only in the sense of deeming it in retrospect to have been the right thing to do, but it cannot count as the actor’s reason for so acting. If to act rationally is to act for a reason, because prompted or motivated by that reason, then action reasoned for only by hindsight, for which reasons are assigned only after the event, falls short of full-blooded rationality.

The problem that animates this study arises when commitment to the prospective nature and inherent norm-governed-ness of rationality comes to bear in considering the possibility of rational norm revision. No
such problem arises for those who deny Normative Diversity and/or the very possibility of rationally contrived normative realignment. But for those of us who do not, the problem of rational norm revision, as we shall explain immediately, is a major one, deemed by most insurmountable.

But first two points of clarification. First, in what follows normative diversity, that is, the plurality of incompatible normative outlooks, will be assumed, not proven or even justified. We take normativity, the self-committed, and commitment-driven nature of human agency, to be a human universal, in the sense of being definitive of sapience. And we join Taylor, Brandom, and others in viewing normativity further as grounded in our discursive capacities. But, bracketing contingently shared areas of “minimal code,” we take no normative content to be fixed or universal.

The second point concerns the possibility of rational norm revision. It is important to understand that the aim of this study is not to prove that framework transitions can be genuinely motivated by rational argument and deliberation. We know that they can. The problem we set ourselves is primarily not that of convincing our readers that norm revision can be rational but that a compelling philosophical account of their rationality is forthcoming—passé Rorty—from the broadly neo-Kantian and neo-pragmatist perspective from which we write. Of all the thinkers whose work is engaged in what follows, Friedman’s Dynamics of Reason steers closest to us in this respect. Friedman does not ask whether framework transitions in physics are rational but rather how to make philosophical sense of their rationality from the essentially Kuhnian perspective to which he is committed. Like Friedman’s, our work too is primarily a work of philosophy.

**The Problem**

Revisable normative frameworks are adjusted and changed in a variety of ways, some surreptitious and unwitting, some contrived and deliberate. Concentrating on the rationality of framework revision, we focus naturally on a subset of the latter category: namely, cases in which a functioning framework is modified or replaced for a reason. Now, the only prospective reason practitioners might have for wanting to alter or replace their
framework is to find it in some sense sufficiently wanting to be reconsidered (as opposed to cases, of which Rorty makes much, in which replacing the old framework is justified in retrospect from the now-perceived superior perspective of the new).

There are three senses in which a functioning normative framework can, in principle, be faulted from within: it could be found incoherent, in some sense lacking, or in some sense wrong. The first two are relatively unproblematic because they do not involve deeming any part of it to be normatively flawed. They are also, as we shall see, largely uninteresting.

Formal failings—inconsistency, incoherence, or insufficient clarity—may well require modifying or even relinquishing lower-level commitments in the light of higher-level standards. But doing so proves the framework to be considered right, not wrong! All the more so in cases where, in engaging other cultures, one’s framework is supplemented by unforeseen normative possibilities encountered. Here, too, amendment is achieved without any part of the original framework being normatively condemned. In the first case, lower-level commitments may well be relinquished, but they will be dropped because they are found expendable, not because they are deemed normatively wanting. In the second, a commitment is added but none renounced.

Both cases obviously represent rational forms of normative adjustment. But they are powerless to account for the rationality of transitions between incommensurable or normatively incompatible alternatives, cases in which practitioners contrive to locate or to fashion frameworks significantly different from their own—as in the case of scientific revolutions. Troubleshooting for inconsistency or incoherence can result in normative reprioritizing within a given system, and in extreme cases, in pruning certain of its expendable elements. Awakening to new possibilities can at most result in supplementing the existing system (but never, as we shall argue at length, in rationally motivating replacing one of the system’s existing elements, which would be in blatant violation of Comparative Irrealism). It is, therefore, to the third option that we direct our attention—where frameworks are revised or replaced, because deemed normatively wanting, an option that would seem to characterize processes of rational norm revision far better than

© 2011 University of Notre Dame
the other two, yet is considered notoriously problematic to the point of incoherence.

Framework transitions to normatively incompatible alternatives, involve, by definition, substituting major elements of the system by others—which to be deemed rational must be undertaken for a reason. Here, in particular, the difference between prospective and retrospective justification is crucial, because the norms governing the reasoning are replaced in the course of the transition. Finding reason retrospectively for having relinquished norms or standards to which one is no longer committed is a matter quite different from finding normative reason to do so while they still bind one.

Hence the problem. To insist that rationality be both prospective and norm-governed and that norm revision can be rational is hence to insist that rational agents not only exercise their norms and standards in self-criticism but also make them the object of such criticism, that one can coherently deem wrong one’s very standards of right and wrong! No viable philosophical account of personal identity makes room for such feats of self-negation. On the other hand, no amount of mere prioritizing or supplementation, we insist, can get one rationally from Aristotelian to Newtonian physics, for example. It follows that if troubleshooting (prospectively) for coherence or (retrospectively) for lacunae is held to exhaust the range of rational normative revision, then Rorty must be granted his point that the classical examples of framework replacement in science and philosophy, or the type of deep social reform explored by Walzer, can simply not be considered rational—a conclusion we firmly contest.

No work to date has addressed this dilemma openly and systematically. This is our aim in what follows. But we do so with a bias. If one is committed, as we are, to the possibility of rational norm revision (pace Rorty), one has one of two options:

(a) to argue for a viable philosophical account of normative self-criticism capable of meeting the obvious objections, or
(b) to argue that the rationality of radical framework transition can be fully and viably accounted for without resort to normative self-criticism, on the basis solely of troubleshooting for normative coherence, and openness to noncritical normative supplementation.
Many, including Brandom, Friedman, and McDowell, whose work we discuss and build on, opt for (b). The bias of the present study is firmly in favor of (a).

The Argument

The remainder of this book consists of three parts. Part I, “Through Thick and Thin,” purports rigorously to make the case for Comparative Normative Irrealism: the principle, to which this study remains committed throughout, according to which two sufficiently diverse normative outlooks are normatively unrankable. Chapters 2 and 3 argue for Comparative Irrealism on semantic grounds, from the different perspectives of two major theories of meaning that dominate approaches that take seriously the neo-Kantian idea of the constitutive role of language. By utilizing the distinction introduced by Bernard Williams and Michael Walzer between thick and thin normative concepts, we are able to chart a robust framework-dependent notion of normative realism, as well as to make a compelling case for Normative Diversity with reference to the former (e.g., concepts such as modesty, valor, or holiness). The idea is to show that with respect to both theories — those that ground meaning in communal norms and idiolect-based theories late of Davidson — the very assumption of Normative Diversity entails Comparative Irrealism, that is, the unrankability in principle of sufficiently alien normative outlooks. The upshot of Part I is, therefore, that if normative self-criticism is at all possible, it cannot be comparative.

The three chapters in Part II, “Rationality from Within,” assess and engage three important latter-day philosophical positions, all of which share premises significantly akin to our own and bear directly on issues central to the question we raise.

Normative Diversity, the incomparability in principle of sufficiently different normative outlooks, and the transformative role of “connected” normative criticism from within are the very principles on which the so-called interpretive approach to ethics is built. Chapter 4 purports, first, to substantiate the interpretive account (in its Walzerian rather than Taylorian form) in terms of the thick/thin distinction developed in Part I,
and in a manner applicable to normativity in general. The second and more critical objective of chapter 4 is to expose and analyze the constraints interpretivism seems inevitably to impose on the kind of normative self-criticism it seeks to promote, which we ultimately find too restrictive. Under an interpretivist construal, normative criticism can venture no further than to challenge other articulations of a given way of life and is powerless to challenge the way of life itself.

Friedman’s work on the rationality of framework transitions in science is the topic of chapter 5. Though limited to science, Friedman’s Kuhnian point of departure steers closest to the presuppositions that animate the present study: forcefully conceding the framework-dependency of science late of Kant, as well as the Normative Diversity and Comparative Irrealism of a science’s successive frameworks—at least to a certain extent. Also appealing is his willingness to deem rational framework transitions exceedingly more radical than those allowed by interpretivists. And yet, as we argue in some detail, unlike interpretivists, Friedman’s account succeeds in steering wide of any mention of the problem of normative self-criticism by focusing exclusively on the so-called problem of incommensurability, namely, that of explaining how, once formed, a new scientific framework can be considered retrospectively a “live option” by practitioners of the old. Lacking from Friedman’s picture is any attempt to account prospectively for the rational incentive to seek or form an alternative to a functioning scientific framework in the first place—without which, we argue, no account of the rationality of scientific paradigm shifts can be considered complete. To do so, we insist, the problem of normative self-criticism is unavoidable.

Chapter 6 is devoted to Robert Brandom’s inferentialist normative pragmatism (appropriately extended from conceptual norms to normativity in general, as suggested by the work of Jeffrey Stout, as well as by some of Brandom’s own more recent writing), in which, we argue, the interpretivist position receives powerful (if unacknowledged) articulation. Brandom’s system offers a rich account of rational action as reasoned action that, not unlike Friedman’s, appears to make no room for criticism. But unlike Friedman’s account, we find in Brandom’s “deontic scorekeeping” (especially when its Kantian and Hegelian underpinnings are sufficiently exposed) an articulation of reasoned discourse that tacitly
attributes to its participants an essentially self-critical motivational
stance. Making Brandom explicit in this regard, so to speak, paves the
way for the constructive account of normative self-criticism we propose
and develop in Part III.

In chapter 7, the first of the three concluding chapters of the book,
we propose, in outline, an essentially pragmatist phenomenology of
(prudent) criticizing as a discursive move, or speech-act, clearly distin-
guished from doubting, questioning, testing, and being merely pressed for
one's reasons. Criticism is analyzed as an addressed act of speech, directed
to alerting its addressees to the existence of a problem or shortcoming
within the compass of their responsibility and, in doing so, prompting
them to take action. When properly distinguished from, say, mere test-
ing, all criticism is shown to contain an element of rebuke. To be criti-
cized is to be held, to some extent, responsible for the shortcomings ex-
posed; to be prompted not only to attend to those shortcomings, but in
some sense also to mend one's ways. Hence, all criticism, it is argued, com-
prises some measure of normative criticism. And because the aim of all
prudent criticism is to convince its addressee to take action, to achieve its
desired transformative effect, it must be endorsed by its addressee as self-
criticism. Normative self-criticism is, therefore, at least to some degree, of
the very nature of all criticism.

Placing the onus of rationality on criticism, as we do, ultimately lo-
cates the transformative locus of rational reckoning in the realm of in-
trapersonal deliberation rather than interpersonal discourse. Participation
in the public game of giving and asking for reasons is certainly the
sure sign of rational engagement. But the outward critical scrutiny of one
another's reasoning will have its rational transformative effect only when
accompanied by a parallel, resonating self-scrutinizing of one's commit-
ments and entitlements. Indispensable as participation in the public game
of giving and asking for reasons is to demonstrating one's rationality, it is
only in the intrasubjective processes of normative self-scrutinizing that
rationality can properly be said to reside and assert itself.

In chapter 8 we therefore turn our attention inward. Focusing on
the so-called hierarchical account of personhood, identity, and norma-
tivity developed by Charles Taylor, Christine Korsgaard, and especially
Harry Frankfurt, the chapter explores the prospects and especially the
limitations it imposes on a person’s capacity for normative self-criticism. We find Frankfurt’s intrapersonal, volitional account of human agency to resonate well and to an extent to reflect, if implicitly, central aspects of the neo-Kantian, pragmatist, and interpretivist pictures of interpersonal normative discourse explored in Part II.

The problem is that no account of human selfhood will suffice of itself to adequately account for the possibility of genuine normative self-criticism—not even Frankfurt’s—for the simple reason that left to her own devices, a person is indeed incapable of taking genuinely critical normative stock of her own normative commitments. The central claim of chapter 8, which is the central claim of the entire study, is that a viably philosophical account of normative self-criticism does indeed present itself but only when the two pictures are amalgamated: when a detailed Frankfurtian picture of the reflective self is grafted onto an equally detailed picture of the kind of critical discursive environment pictured by Brandom and Walzer. This has so far never been seriously attempted. Those, like Taylor, Korsgaard, and Frankfurt, whose interest lies in the kind of reflective, intrasubjective self-reckoning central to questions of self, agency, and practical reasoning, have given little if any thought to the possible bearings on it of the external intersubjective discursive contexts in which the selves they study constantly partake. Conversely, thinkers like Brandom, McDowell, and Walzer, whose interests lie in the social, dialogical arenas of communal critical discourse, show no interest in following through their possible transformative impact on the personal commitments and self-reasoning of individual discussants. Intrasubjective reflection and intersubjective discourse are studied by philosophers quite independently, as if the extent and quality of a person’s dialogical dealings had no bearing on his reflective capacities, and vice versa.

Combining the two realms of discourse does not, however, yield as straightforward a solution to the problem of normative self-criticism as one might expect. For the very same reason that it is impossible to argue normatively against the norms to which one is committed, it is impossible to be convinced to do so by others. If normative self-criticism is unthinkable, it remains unthinkable when leveled by others. But if normative criticism is incapable of convincing in what sense can it be said to be rationally internalized and endorsed?
The answer proposed in chapter 8 follows on the heels of the conclusions of chapter 7. For their criticism to be endorsable by those they criticize, we argue there, prudent critics must base their case on premises their addressees hold true. But in the case of normative criticism, this is never an option. No premises exist to which a person is liable to agree, that entail a denunciation of his very norms. Sensing this, normative critics tend to argue from premises close to what their addressees hold true, yet sufficiently different to allow their arguments to stick. (Arguing from the left, critics will surreptitiously premise certain liberal norms to make their case, whereas arguing from the right, they will tend to smuggle in just enough conservative value to prove their point.) Normative criticism thus typically conveys a portrayal of its addressees’ relevant commitments that differs minimally yet crucially from their own self-image. Because normative criticism challenges heartfelt norms, it is powerless to convince, but if trusted, we argue at length, the discrepancy between the two portrayals, as in the case of a disturbing playback (an analogy explored in some detail), is capable of undermining its addressees’ initial commitment and of rendering them ambivalent toward those norms. Herein, we argue, lies the destabilizing, or “ambivalating,” potential of trusted normative criticism. And, as Frankfurt famously argues, norms to which one becomes ambivalent lose their wholehearted volitional grounding and normative hold and demand to be reassessed.

Chapter 9 takes stock of the book’s project by revisiting the dynamics of scientific framework transitions in light of the conclusions of chapter 8. If, as McDowell puts it, an ability to “reflect about and criticize the standards by which, at any time, [we take ourselves] to be governed … from the midst of the way of thinking [we are] reflecting about,”\(^{58}\) becomes a philosophically articulable option only when a critically motivated version of Brandom’s interpersonal deontic interrogation is allowed to bear on a Frankfurtian normative dialogue of self (or some similarly hierarchical equivalent), then, to be considered rational, Friedman’s Kuhnian picture paradigm shifts require substantial supplementation on both counts. Friedman, as we argue in chapter 5, reduces the problem of rational framework replacement in science to that of communicating rationally across an already existing divide. We take it primarily to be that of the rational incentive to create a new framework in the
first place. To answer this question, one cannot avoid asking what fault such a person could have found in the framework to which he was still committed—which in light of the conclusions of chapter 8, we argue, can now be addressed from within Friedman’s approach, in ways that have so far been unavailable from that perspective.

It is not an easy fit, however. As we have seen, the sort of self-critical disposition needed for setting a paradigm shift in motion requires the challenging environment of trusted, potentially ambivalating criticism. With all due respect to the importance Walzer assigns to the “connectedness” of effective critics, from the neo-Kantian perspective we share with Friedman, such criticism will clearly not be forthcoming from within the “normal,” paradigm-governed discursive settings of one’s own field. The initial source of destabilization has to be external. On the other hand, a scientific field can only be transformed from within. Only physicists can change physics! Accordingly, chapter 9 proposes in rough outline a two-stage format for studying scientific paradigm shifts, in which practitioners of voice and standing are first ambivalated by external critics and later (inadvertently) succeed in promulgating their newfound indecision within their communities by means of a particular form of published work. Both stages represent irreducible moments of intense intrasubjective deliberation, which do not admit easily to the collectivist modes of explanation that have come to dominate science studies since Kuhn.

To account for the first stage, we propose an extension and reworking of Peter Galison’s novel notion of the scientific “trading zone”—the “locales” of professional engagement outside one’s home community, as when bidding for financial support, “trading” with neighbor disciplines for techniques or instruments, offering scientific opinion, engaging students, and so forth. The point we stress is how when “trading” abroad one is frequently exposed to the friendly and trustedly bemused questioning of genuinely curious professionals, on whose thinking the framework constitutive of one’s own may have far less of a grip. Here, we show, lies the destabilizing and ambivalating potential of science’s external critics.

The second stage is achieved when the doubts and indecision of duly ambivalated individuals come tacitly to inform their creative efforts to overcome them. Such efforts are shown to typically take the form of uneasily split, hybrid attempts to re-represent their field’s basic assumptions
by retaining some of the old while groping imaginatively toward other possibilities. When analyzed prospectively, such works can be seen to represent anxious yet highly creative departures from heartfelt commitments capable of motivating others to seek cleaner, more radical breaks with the old. Classical examples we mention briefly are Tycho’s planetary theory, Galileo’s analysis of projectile motion, and Poincaré’s geometrical conventionalism, all of which we show unwittingly preserved, and therefore propagated, the keen ambivalence that begot them, prompting others to take a firmer stand.

The philosophical problem this study raises and purports to address—that of articulating the possibility of normative self-criticism from a broadly neo-Kantian perspective—is deemed by those who deny Normative Diversity to be misconceived, others (Popper’s school, for instance) deem it trivial, but most (explicitly or implicitly) hold it to be insurmountable. Because we concede Normative Diversity and refuse to limit self-criticism to instrumental reason, in what follows the work of the first two groups is largely ignored. We contest the third group’s conclusions but not by accusing them of misconceiving the problem or of reasoning it through mistakenly. For we fully concede their claim (or assumption, or tacit presupposition) that, of her own accord, a person is indeed incapable of mounting a normative argument against her own norms and, consequently, of being convinced to do so by others. If we accuse them of anything, it is of oversight, of failing to notice how the destabilizing impact of rational discourse can extend rationally beyond the realm of reasoned exchange. This is the modest undertaking of the chapters that follow.