On Reason and Rationality

I've written a number of interrelated papers:
1. "Why Be Rational?," *Mind* 114:455 (2005): 509–63.
2. "State or Process Requirements?," *Mind* 116:462 (2007): 371–85.
3. "Reply to Bridges," *Mind* 118:470 (2009): 369–376.
4. "How Does Coherence Matter?," *Proceedings of the Aristotelian Society* 107:1 (2007): 229–263.
5. "Why Be Disposed to Be Coherent?," *Ethics* 118:3 (2008): 437–463.
6. "The Myth of Practical Consistency," *European Journal of Philosophy* 16:3 (2008): 366–402.
7. "Toward a Principle of Instrumental Transmission"
8. "Ought: Between Subjective and Objective," co-authored with John MacFarlane
9. "Ought: Between Justification and Truth," based on joint work with John MacFarlane
10. "Aims as Reasons," in *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon*, edited by Samuel Freeman, Rahul Kumar, and R. Jay Wallace (New York: Oxford University Press, forthcoming)

But these papers are awfully convoluted themselves, and certainly, the relations between them are hard to follow. To make matters worse, my thinking has changed along the way. So here's an attempt at an overview of how I now see these matters.

No doubt, 'ought' and 'reason' are used in lots of ways. I'm interested in a distinctive use: that of guiding first-person deliberation about what to believe, feel, intend, or do—and, by extension, that of offering advice, which tries to help someone else reach a deliberative conclusion. To my mind,

> In deliberating about what to believe, what matters is just what is likely to be, or what the evidence suggests, is true.

> In deliberating about what to intend (or do), what matters is just what is likely to, or what given the evidence will, promote, respect, etc. what is of value.

So I find plausible:

> *Theoretical*: One ought to believe that P iff it is sufficiently[1] likely that P is true.[2]

> *Practical*: One has reason to intend (or do) something to the extent that it is likely to promote, respect, etc. what is of value.[3] What one ought and ought not intend (or do) is a matter of one's reason to intend (or do) it relative to the relevant alternatives.

---

[1] This may depend on what is practically at stake.

[2] Maybe better: One ought to believe that P if it is sufficiently likely that (P is true *and would still be true if one were to believe it*) and *one is, or ought to be, considering whether P*. One ought not to believe that P if it is not sufficiently likely that (P is true and *would still be true if one were to believe it*).

Why not the following "objective" alternative to Theoretical, which drops the "likely" bit?

> One ought to believe that P iff P is true.[4]

Because of cases like this: Suppose a fair coin has landed, but I don't know how. I would think, "I ought not believe that it landed heads, and I ought not believe that it landed tails." But if I ought to believe just what is true, then I should know that what I have said is false: either I ought to believe heads and I ought not believe tails, or I ought to believe tails and I ought not believe heads. This is why Theoretical is sensitive to the probabilities, not what is actually so.

The same goes for the following alternative to Practical:

> One has reason to intend (or do) something to the extent that *will* promote, respect, etc. what is of value.

Suppose ten miners are trapped in one of two shafts, but I don't know which. If I block A, all ten are saved if they are in A, all die if in B. If I block B, the reverse. If I block neither, nine are saved wherever they are, but one dies. I would think, "I ought to block neither." But if I ought to do just what *will* promote value, then I should know that what I have said is false: either I ought to block A, or I ought to block B. This is why Practical is sensitive to the probabilities, not to what is actually so.

Beyond this, however, I don't try to *defend* Theoretical and Practical, (although of course how *interested* you will be in what I go on to say will depend on whether you also find them plausible too!). My aim is instead to describe a puzzle that arises *if we accept them*. The puzzle is that there seem to be requirements—which I'll call, a little artificially, "rational requirements"—that we believe or intend *what coheres with our attitudes or evidence*. And they tell us to believe or intend what is often *something other than* what is likely to be true, or to promote, etc. value.

These "rational requirements" come in three varieties. (The important point for me is simply that it is often thought that there are requirements with these contents. Whether all or any are properly called "rational" is of less concern.)

1. *Formal coherence requirements*: that we ensure formal—logical or arithmetical—relationships among the contents of our beliefs and intentions:
   - N: one is required not (to believe P and to believe not-P)
   - C: one is required not (to believe P, but to fail to believe Q) where Q is a logical consequence of P.
   - ME: one is required not (to intend E, to believe that one E's only if one M's, and to fail to intend M)

---

[3] This may depend on our individual relations to things of value (e.g., my friendships may give me reasons that yours does not).

[4] Or perhaps: P *is true* and would still be true if one were to believe it, and one is, or ought to be, considering whether P. One ought not to believe that P if P *is false*.

- IC: one is required not (to intend E1, to intend E2, and to believe that if one E1's, one does not E2)

Note that "required" takes "wide scope," since coherence can be achieved in more than one way; there is no particular attitude that one must have or lack in order to satisfy the requirement. Note also that various refinements and qualifications are possible.

2. *Kratic requirements*: that we comply with the contents of our own normative beliefs
- K+: If you believe that you ought to have belief or intention A, then you are required to have A.
- K-: If you believe that you ought not to have A, then you are required not to have A.

In my opinion, these requirements are best understood as narrow scope (and as governing processes, rather than states), but this is more controversial. Again, various refinements and qualifications are possible.

3. *Subject's-evidence requirements*: that we believe or intend in accord with what is likely *given our evidence*
- If P is not sufficiently likely *given your evidence*—or if you would not be *justified* in believing P—then you are required not to believe P.
- If X is something you ought to do *given your evidence*—of if you would not be *reasonable* in failing to X—then you are required to X.

I take these too to be narrow scope; one doesn't satisfy them by changing the evidence. (Henny Youngman: "When I read that drinking was bad for my health, I gave up reading.")

It is also worth noting in this connection:

4. *Desire-, or will-, based theories of our ultimate practical ends*. In place of Practical, these say:

> One has reason to intend (or do) something to the extent that it is likely to promote, respect, etc. *what we desire or will*.

These theories seem different from 1–3. They don't put forward requirements that are somehow supposed to coexist with Practical. Instead, they claim that the practical end is not value, but instead what the agent desires or wills. So I think they call for different treatment.

In some earlier work, I drew this (admittedly artificial) distinction between reason and rationality by saying that: whereas the reason that there is for us to have certain attitudes *depends only on facts,* rational requirements are such that whether one satisfies them *depends only on our attitudes,* in abstraction from the facts. I now think that this was inapt, because:
1. The arguments for the factivity of reason look weaker when one accepts the relativist account of 'likely'.
2. If evidence is factive, then subject-evidence requirements fall on the wrong side of the line.

Formal coherence and kratic requirements:
Take N as our example. Recall that N says that I am required (either not to believe P or not to believe not-P). But Theoretical says that I ought to believe just what is sufficiently likely to be

true. How are these compatible? Suppose P is sufficiently likely to be true and not-P isn't. And suppose that I drop P, but keep not-P. Then I have satisfied N, but I certainly haven't followed Theoretical.

1. Patterns of reasons:
Here's a quick error theory for N. The only reason N seems plausible to us, this error theory claims, is the:

Violation Intuition: If one believes P and believes not-P, one violates some requirement.
But to explain VI, we don't need:

N: one is required (either not to believe P or not to believe not-P).
We can instead appeal to the following:

First fact about reason patterns: In any given case, either one is required by reason (i.e. ought) not to believe P, or one is required by reason not to believe not-P
(Note the change in scope; it's a disjunction of requirements, rather than a requirement of a disjunction.) First Fact is consistent with Theoretical.

Why explains the patterns of reasons associated with…
1. … N?
   - *Stronger evidence*: One ought to believe P only if P is more likely than not-P.
   - Perhaps *SE* is in turn explained by the fact that avoiding false belief is more important than acquiring true belief.

2. … C?
   - Why is it the case that either I ought not believe P, or I ought to believe Q?
   - Suppose it is *not* the case that I ought not believe P. Then by…
   - *Epistemic Strictness*: Either it is the case that one ought not believe P, or it is the case that one ought to believe P
   - … I ought to believe P.
   - *Evidential Transmission*: If Q must be the case if P, then Q is at least as likely as P.[5]
   - Hence, if I ought to believe P, then I ought to believe Q.

3. … IC?
   - Why is it the case that I have *less* reason (to intend to take the job at Rutgers and to intend to teach 12A at Berkeley) than I have for *some alternative* pattern of intentions?
   - *General Transmission*: If there is reason for one to *E*, then M-ing inherits this reason in proportion to the probability that, conditional on one's M-ing, (one *E*'s and one's *M*-ing helps to bring about one's *E*-ing in a nonsuperfluous way).
   - *Intentions are means*: "M-ing" here can be "*intending* to E."
   - *Predictive significance of intention*: Intending to E1 (e.g., take the job at Rutgers) typically makes it more likely that one will E1. The fact that it is more likely that one will E1 may affect what one is likely to achieve by intending to E2 (e.g., teach 12A at Berkeley).

---

[5] I'm suppressing some complications here.

- Via General Transmission, that affects one's reason to intend to E2. (If you plan on taking the job at Rutgers and are so likely to take it, there's no point in planning to teach 12A here at Berkeley, since it will be fruitless.)

4. … ME? Similar to IC.

5. … K+ and K-?
   - Why is it the case that either I ought not believe that I ought to X, or I ought to X?
   - Suppose it is not the case that I ought not believe that I ought to X. Then by *Epistemic Strictness*, I ought to believe that I ought to X.
   - If I ought to believe that I ought to X, then by *Theoretical* it is sufficiently likely that I ought to X. Then by *Practical*, it is sufficiently likely that it is sufficiently likely that X-ing responds to what is of value etc.
   - By a *probability-collapse principle* (e.g. If it is likely that it is likely that P, then it is likely that P), then, by Practical, again I ought to X.

Shortcomings of this quick error theory:
1. Are these "Facts" plausible? In particular, Epistemic Strictness?
2. The First Fact doesn't explain the:
   > *Satisfaction Intuition*: Suppose one believes P and believes not P. If one drops one's belief that P and keeps one's belief that not P, even though reason requires one to believe P and not to believe not P, one satisfies a requirement, or gets something right, in a way that one would not have if one had continued believing P and believing not P.

   However, N does explain the Satisfaction Intuition.

2. Some failed attempts to reconcile N with Theoretical
"Suppose one ought not believe not-P. A necessary condition of not believing not-P is (either not believing P or not believing not-P). One ought to fulfill any necessary condition of what one ought to do. So one ought (either not believe P or not believe not-P). Of course, one hasn't satisfied a *sufficient* condition of what one ought. But one has satisfied a *necessary* condition."
1. First, I think it's false that I ought to satisfy any necessary condition of what I ought to do.
   o Frank Jackson's Professor Procrastinate, who must agree to write the review in order to write it, but won't write the review even if he agrees to write it. Although he ought (to agree and write)—which would bring it about that the review got written—he ought not (to agree)—which will bring it about that the review doesn't get written. (General Transmission is meant to be sensitive to such cases.)
   o Ross's paradox. (See Fabrizio Cariani's dissertation)
2. Second, by the same logic: "Since I ought not believe not-P, I ought (not believe P or not believe not-P *or both believe P and believe not-P*). By both believing P and believing not-P, I satisfy a requirement in the same way as by not believing P." But the Satisfaction Intuition was that one satisfies a requirement, or gets something right, *in a way that one would not have if one had continued believing P and believing not P*.

"By satisfying N over the long run, one better satisfies Theoretical over the long run."
> 1. Even if this is true about the long run, in what sense do I do something I ought *in this particular case*?
> 2. By satisfying *formal* coherence requirements, one does *not* better satisfy Theoretical over the long run. (Kratic requirements are different, provided that one's normative beliefs tend to be reliable.)

"There just is some intrinsic value in satisfying N."
> 1. Implausible, as though psychic tidiness were an end in itself.
> 2. This intrinsic value would have to be practical. So as far as belief is concerned (as it is with N), this still wouldn't satisfy Theoretical.

3. Evaluative, not deliberative:
Perhaps these requirements aren't *meant* to play any direct role in deliberation and advice, in which case there would be no conflict with Theoretical and Practical. They might instead be merely evaluative standards, which we use to appraise performances, as bystanders might.
> "The heart *ought* to pump blood, and it is pumping blood. So it is *functioning well*."
> "That was a *skillfully performed* murder."

Note:
> 1. Evaluative standards *sometimes* play a role in the deliberation of the *agent* whose performances are being evaluated. ("I ought to turn in that way, because it would be graceful.") The point is that we can make sense of evaluative standards apart from any such role.
> 2. Perhaps evaluative standards *always* play a role in the deliberation of *appraisers*. Perhaps: "… is skillful" implies "… gives appraisers reasons to admire it." But the appraisers need not be the agent.

Shortcomings:
> 1. In what way is satisfying a *formal coherence* requirement *functioning well*, etc.? As we saw earlier, we can't answer: "Because it is manifesting a disposition that is good to have, insofar as that disposition leads one to better satisfy Theoretical and Practical over the long run."
> 2. Kratic requirements (at least) *seem* to play a role in deliberation and advice that this doesn't account for.

4. Transparency:
Consider K+:
> 1. K+ can *seem* to have significance for first-person deliberation: The subject believes, of the response that is required, that she ought to give it.
> 2. K+ can *seem* to have significance for second-person advice: The adviser informs the subject of what is required, with the expectation that, as a result of this advice, the subject will give the required response.

But this can be explained simply by the *content* of K+.

Shortcoming: Works only for kratic, not formal coherence, requirements.

5. Beliefs about patterns of reasons:
Given general awareness of what one believes and intends, and given the availability of the facts about patterns of reasons, the following seems plausible:
> *Self-monitoring*: if one violates a formal coherence requirement, one ought to arrive at some belief about which of the relevant attitudes, if any, are at odds with reason.

Now, suppose that one violates a formal coherence requirement. Then:
   i.   One arrives at some such belief about which of the relevant attitudes, if any, are at odds with reason.
   ii.   One revises one's attitudes in accordance with that belief.

If one satisfies a formal coherence requirement in *this* way, then, even if one satisfies it in the *wrong* way, one still satisfies:
   i. Self-monitoring, which is a requirement of reason.
   ii. K+ and/or K-, whose seeming deliberative/advisory role we explained under 4 above.
Perhaps this explains the sense that when one satisfies a formal requirement, even if one satisfies it against reason, one does something that one ought to do.

Shortcoming: What if one does *not* satisfy N in this way? Reply: But is it so clear that one does get something right? Perhaps the Satisfaction Intuition is simply overstated.

Subject's-evidence requirements:
"Why is there even a problem? The subject's-evidence requirements are *identical* with Theoretical and Practical! If *your* evidence supports P, then P *is* likely."

1. Transparency (in a different sense!)
In deliberating about what to believe, we don't think about ourselves, but how things are. But on this view, we are ultimately always thinking about our own condition: what evidence *we* have.

2. Conditional contexts
In conditional contexts, justification and ought/likely come apart.
> "If the butler had had motive, but my evidence indicated that he did not, it would still be likely that he was guilty, but I would not be justified in believing it."

> From Yalcin:
> #"If not P but it's likely that P, then…"
> "If not P, but I'm justified in believing P, then…"

3. Advice
When the speaker is not the believer, justification and ought/likely come apart:
> Patient: "Ought I to believe that I have cancer? That's what the last specialist said."

> #Doctor: "Well, then, you obviously ought to believe it. Why are you asking me?"

Doctor: "You're certainly justified in believing that; that specialist has a fine track record. But no, thankfully it's not what you ought to believe. It's not likely that you have cancer; that specialist overlooked something."

Nevertheless, we can explain why the subject's-evidence requirements are plausible.

1. First, they are genuine *evaluative* requirements. When we're not deliberating or advising, but appraising someone's performance as a believer or agent—whether she's responsible, someone to emulate, etc.—what matters is how well they have responded to *their* evidence, not to ours.

2. Second, they can *seem* like genuine *deliberative* requirements, because they coincide with Theoretical and Practical *in the first person*. (Compare the kratic requirements and Transparency, in the first sense.)

   #"My evidence supports P, I'm justified in believing P, but P isn't at all likely."

What can "likely" or "the evidence" mean here? Surely the "P is likely" is true or false relative to *some* body of information. Two questions:

1. *Which* body of information is relevant for evaluating the truth or falsity of the proposition expressed by an occurrence of "P is likely"?

   A natural thought is that it is the *speaker's* body of information (more precisely, the information relevant at the context of occurrence). But then it would be hard to make sense of disagreement, retraction, and advice. Why aren't patient and doctor simply talking past one another?

   Patient: "It's likely that I have cancer. That's what the last specialist said."
   Doctor: "No, it isn't likely. He made a mistake."
   Patient: "OK. I take that back."

   Patient: "Given my evidence now, it's likely that I have cancer."
   #Doctor: "No, given my evidence, it's not likely that you have cancer."
   #Patient: "OK. I take that back. Given my evidence then, it wasn't likely."

   Compare:
   Patient: "I am not a doctor."
   #Doctor: "No, I am a doctor."
   #Patient: "OK. I take that back. I am a doctor."

   Following John MacFarlane, I think the relevant body of information is the *assessor's* (more precisely, the information relevant at the context of assessment). Relative to the doctor's information, what the patient said was false.

2. Is the proposition expressed by an occurrence of "P is likely" *about* someone's information, e.g., the assessor's?

   No: the conditional-context examples show why.

Compare the proposition expressed by an occurrence of "Obama is President." If this were a claim *about the actual world*, then it would be necessarily true. (It is true at every world that at *this* world, Obama is President.) Rather, the same proposition (function from worlds to truth values) is true at this world, but false at other worlds.

Likewise, the same proposition expressed by "P is likely" is *true or false* relative to a body of information, but isn't *about* one.

Implications for the factivity of reasons and evidence:
Weird:
> # "(i) There is reason to believe P, namely Q. (ii) But Q is not the case."

One response: This means:
> # "(i*) I would be justified in believing P because I believe Q. (ii) But Q is not the case."
and that's Moore's Paradox.

But we've shown that (i) doesn't mean (i*). An *assessor* will assess (i) and (ii) as contradictory, but need not assess (i*) and (ii) as contradictory. Are we to conclude, then, that reasons *must be* facts?

Not necessarily. Suppose that
   a. "There is reason to believe P, namely Q" is true relative to a body of information only if Q belongs to that information.
   b. My information is *everything I believe, whether or not it is true*.
   c. I believe Q.
   d. Q is not the case.
Then relative to my information, (i) is *true* only if I believe Q. And if I believe Q, then I will not *accept* (ii). Thus, while (i) and (ii) can be (and in this case are) both *true* relative to the relevant body of information and world, it can't be the case that (i) is true relative to my information and (ii) is acceptable to me. In sum, it will always *look* like reasons are factive. But that doesn't mean that they are.