

## Lectures on Reasons and Rationality

Niko Kolodny, UC Berkeley

[kolodny@berkeley.edu](mailto:kolodny@berkeley.edu)

Papers referenced may be found here: <http://sophos.berkeley.edu/kolodny/papers.htm>

### Overview

1. Normative reasons as what matter in deliberation and advice
2. Normative reasons related to ought.
3. What the agent has reason, or ought, to do is neither “objective”—depending on all the facts—nor “subjective”—depending only on the agent’s evidence.
4. Instead, reason and ought depend on what *epistemic modal* judgments depend on: what is likely, what must be the case, etc. And that is the information of the *assessor* of the relevant claim.
5. One implication of this view is that modus ponens is invalid for the indicative conditional.
6. Another implication of this view is to undercut arguments that reasons are facts.
7. Rational requirements, as depending on how things are with the subject: (i) formal coherence requirements, (ii) kratic requirements, (iii) subjective-evidence requirements.
8. A puzzle: What rational requirements tell us to do differs from what, on plausible, substantive views, we have reason for, or ought to, do.
9. An error theory for subjective-evidence requirements.
10. An error theory for kratic requirements.
11. A template for an error theory for formal coherence requirements, focusing on requirement N.
12. How the template would apply to other formal coherence requirements: C, IC, M.

### Normative reasons, as what matter in deliberation and advice

#### What kind of responses can a subject have reason, or reasons, for?

- Responses that are sensitive to the subject’s judgments that there is reason, or are reasons, for such responses (Scanlon, *What We Owe to Each Other*).
- Beliefs, emotions, actions, but not headaches, thirst.

#### Normative, motivating, and explanatory reasons:

1. *Normative* reason: What matters for the subject’s *deliberation* about what to believe, etc.—and, by extension, an advisor’s *advice*, which tries to help the subject reach a deliberative conclusion.
2. *Motivating reasons*: What the agent *took* to be a normative reason. (The taking can be implicit.) “His reason for going to war was that the priest gave an auspicious reading of the entrails.”
3. *Explanatory reasons*: What explains why the agent had an attitude: merely a species of brute cause. “The reason he believes that he is Napoleon is that he was hit on the head.”

#### Reason vs. reasons

1. ‘Reason’ as a *mass* noun: “There is some reason for one to X.” “One has reason to X.” “There is more (stronger, weightier, more stringent, more compelling, etc.) reason for one to X than for one to refrain from X-ing.”
  - Claims that X *has a certain status*.
  - What status? That it is, to some extent, worth having or doing. But that may have misleading connotations. More neutral, but perhaps less helpful: is the status of being supported, in some way or to some degree, by reasons.
  - Perhaps analogous to a preference ordering in decision theory.
2. ‘Reasons’ as a *count* noun: “There is a reason for one to X, namely: that P.” “P is a reason for one to X.”
  - Something that *confers* that status on X-ing, or *in virtue of which* X has that status.
  - The sense of ‘reason’ in which Dietrich and List seem to be interested. This has not had an analogue in decision theory.

One might be tempted to reduce the count-noun sense to the mass-noun sense plus some neutral conception of explanation: i.e., what it is for P to be *a reason* for one to X is for P to help to explain why *there is reason* for one to X. (Compare Broome, “Reasons.”) But I am skeptical. Not only because the notion “there is reason” may just be “is supported by reasons,” but also because not just any kind of explanation will do. Some fact might help to explain why there is reason for one to X without being a (normative) reason for one to X. It might be part of a causal explanation of why that (normative) reason obtains.

Distinguish the mass noun sense from:

3. Reason as a *faculty*: Presumably, the faculty of responding appropriately to reasons.

### **The relation between reason and ought**

‘Ought’ is used in lots of ways. I am focusing on a *deliberative-advisory* use, according to which ‘ought’ is equivalent to “has conclusive normative reason.”

- Contrast: “He ought to be home by now.” “Toasters ought to brown bread.” Neither of these plays a direct role in deliberation and advice.

One ought to X relative to a set of alternative actions/beliefs S if and only if

- i. X is a disjunction of alternatives in S,
- ii. there is *at least as much* reason for one to X as there is for one to perform any alternative in S, and
- iii. *sufficiently more reason* for one to X than there is for any alternative Y that is not a disjunct of X.

(Indebted to Cariani, UC Berkeley Ph.D. Diss.)

Notes:

1. By default, the question “Ought I to X?” has as  $S = \{X, \text{refrain from X-ing}\}$ .
2. What about reason *against* X-ing? Perhaps these are just reasons *to* refrain from X-ing.

Validates:

- If one ought to X, then it is not the case that one has more reason to refrain from X-ing than to X.
- If X and Y are distinct and one ought to X, then one not ought to Y.
  - *Objection*: “Wait: If I ought to mail a letter, then it is not the case that I ought to affix a stamp?!”
  - *Reply*: A context shift, as the question shifts from “Ought I to mail?” to “Ought I to affix?” Relative to {mail, refrain from mailing} ought mail, and relative {affix, refrain from affixing} ought affix.

Does not validate:

1. If there is more reason for one to X than there is for one to refrain from X-ing, then one ought to X. Suppose  $S = \{\text{refrain from X-ing, X, X and Y, ...}\}$ . It might be the case that one refrain from X-ing  $< X < X$  and Y.
2. If it is necessary (logically, epistemically, historically, etc.) that  $X \supset Y$  and ought X, then ought Y. Standard in deontic logic and quantificational semantics for modals (indeed the semantics that Kolodny and MacFarlane, “Ifs and Oughts,” uses for simplicity). But far from intuitive:
  - Ross’s paradox: Ought post the letter implies ought (post the letter or burn it).
  - Frank Jackson’s Professor Procrastinate: 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.
  - Telegraphing boxer: at every possible world at which he throws the punch, he telegraphs it. Ought to throw implies ought to telegraph.

### Reason and ought neither “objective” nor “subjective”:

*On what information* does what we have reason for depend?

- Objectivists say: on *what is actually the case, on all the facts*.
- Subjectivists say: on the *subject’s evidence* (where “evidence” can be understood in lots of different ways).
- I think: Neither. Instead, it depends on what is *likely*, or what *the* evidence supports.
- But on what does *that* depend? I think: on the *assessor’s* evidence.

Note: This is not the same question as: What are *reasons*? Facts, attitudes, contents of attitudes?

1. “The information on which what we have reason for depends” may be broader than “our reasons.” This is because parts of this information may affect the *strength, urgency, etc.* of reasons, without those parts *being* reasons. For example, those parts may affect the probability that a certain value is realized.
2. Even if the information on which what we have reason for depends is just the set of our reasons, the view that all reasons are facts is not the same as objectivism. A subjectivist might take the subject’s evidence to consist, e.g., in what the subject *knows*, and so only in facts.

**Against objectivism:**

*Problem:* The function of reason and ought is to guide deliberation. But deliberation is typically with imperfect information. We often don't know the facts. So objectivism is not the right account of reason and ought.

Moore's Gambit: Of course, we often don't *know* all the facts. But that need not stop us from *judging* that we ought (objectively) to do or believe something, so long as we are *justified* in believing that we ought (objectively) to do or believe it.

Moore's Gambit fails:

*Action example:*

Suppose ten miners are trapped in one of two shafts, but we don't know which.

- If we block A, all ten are saved if they are in A, all die if in B.
- If we block B, all die if they are in A, all ten are saved if in B.
- If I block neither, nine are saved, but one dies, wherever they are.

We would naturally think, "We ought to block neither." But we are *not* justified in believing that we ought objectively to block neither. We *know* that it is false.

*Belief example:*

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 I know that it is false that I ought (objectively) not to believe heads and I ought (objectively) not believe tails.

### **Against subjectivism:**

The lesson, I think, is what I ought, or have reason, to believe or to do depends on whatever it is that *epistemic modal* judgments express: what is *likely* the case, is *possibly* the case, *must* be the case, what *the* evidence suggests is the case, etc.

But it is a mistake to think of what epistemic modal judgments express, and so what I ought or have reason for, as depending on *my* evidence, as subjectivism claims.

The main problem is that this does not make sense of *advice*:

Patient: "Ought I to believe that I have cancer? (Is that likely?) That's what the last specialist told me."

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

Doctor: "No, thankfully it's not what you ought to believe. It's not likely that you have cancer; that specialist overlooked something."

Ignorant rescuer: "Ought I to block A?"

#Foreman, who knows the miners are in A: "No, it's not so that you ought to block A! You don't know where they are!"

Foreman, who knows the miners are in A: "Yes, you ought to block A! That will save all ten!"

Pritchard's Gambit: "You ought to block A" means "It will be the case that you ought (subjectively) to block A, once I transmit my evidence to you."

Pritchard's Gambit fails: The better-informed person may not be able to transmit his evidence. Yet he might still say, or think, what he does in the above dialogues.

*Problem:* Reason and ought not only guide deliberation, but also figure in advice, which aims to aid deliberation. But advisors may have better information than the subject. So reason and ought should not be confined to the subject's evidence.

*Evaluation vs. deliberation*

*Objection:* "But surely Patient would be *justified* in believing is that he has cancer. And surely it would be *reasonable* for Ignorant Rescuer to block neither. Thus, Patient *ought* to believe it, and IR ought to *block* neither."

The premises are correct, but the conclusion does not follow. "Justified/reasonable" and "reason/ought" play different roles. "Justified/reasonable" *evaluate*, whereas "reason/ought" guide *deliberation*.

Crucial to distinguish *evaluation/criticism/assessment* from *deliberation/advice/decision-making/judgment-forming*. We can *measure* someone or something *against a standard* without any direct connection to deliberation or advice.

- "That was a *defily executed* murder." (Not even a *pro tanto* reason to murder.)
- "The heart is pumping blood. So it is *functioning well*."
- "The sunset is *beautiful*." (Hearts and sunsets don't deliberate.)

Note:

1. No doubt, 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 we can make sense of evaluative standards only in terms of the role that they play in the deliberation of *appraisers*. Perhaps: "... is deft" implies "... gives appraisers reasons to admire it." But still the appraiser, who has the reasons to admire, need not be the subject, whose responses are being evaluated.

Two other problems with identifying ought/reason/epistemic modals with the subject's evidence /justified/reasonable:

1. *Transparency*

In deliberating about what we ought to believe or do, we don't think about ourselves, but instead about how things are. But according to subjectivism, we are always reflecting on our own condition: what evidence *we* have.

2. *Conditional contexts*

In conditional contexts, justified/reasonable/the subject's evidence diverges from ought/reason/epistemic modals.

*Subjunctive conditionals:*

I've concluded that the butler had motive, and so that it is likely that he is guilty.

"If the butler still had motive, but my evidence mistakenly indicated that he did not, it would still be likely that he was guilty, although I would not be justified in believing it."

#"If the butler still had motive, but my evidence mistakenly indicated that he did not, it would no longer be likely that he was guilty."

"If the butler did not have motive, but my evidence still indicated that he had, it would no longer be likely that he was guilty, but I would still be justified in believing it."

#"If the butler did not have motive, but my evidence still indicated that he had, it would still be likely that he was guilty."

*Indicative conditionals:*

#"If P, but it's not possible that P, then..."

"If P, but I'm not justified in believing P, then..."

(Indebted to Yalcin, "Epistemic Modals")

### **An alternative**

How then are we to understand ought/reason/epistemic modal judgments? Two questions here:

1. *Which* body of information is relevant for evaluating the truth or falsity of the propositions expressed by occurrences of ought/reason/epistemic modal sentences?

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 and retraction. 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."

*Note:* Disagreement and retraction are *also* problems for the very common suggestion that (say) “ought” is ambiguous, with an “objective” and a “subjective” sense.

With 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 is false. This is why he disagrees with what the patient says. And once the patient is persuaded, then relative to his new information, what he said earlier is false. This is why he retracts it.

2. Are propositions expressed by occurrences of ought/reason/epistemic modal sentences *about* someone’s information (e.g., the speaker’s, the assessor’s)?

No: First, implausible that the deliberator is saying something about the *assessor’s* information. Second, the transparency and the conditional-context arguments suggest that ought/reason/epistemic modal judgments can’t be about someone’s body of information.

But this may seem puzzling. How can the truth of these judgments *depend on* someone’s information without *being about* someone’s information?

Compare the proposition expressed by an occurrence of “Obama is President of the U.S.” 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 an occurrence of “P is likely” (or “One has reason to X” or “One ought to X”) is *true or false* relative to a body of information, but isn’t *about* one.

Think of the proposition that an occurrence of a sentence expresses as a function from worlds *and body of information pairs*  $\langle w, i \rangle$  to truth values. An occurrence of a sentence, S, is true relative to a context of occurrence and context of assessment iff S is true at  $\langle w, i \rangle$  where w is the world of the context of occurrence and i is the information of the context of assessment.

This solves the basic problems with objectivism and subjectivism.

- *Deliberation* is possible, in contrast to *objectivism*. When the deliberator is deliberating, he assesses his own reason/ought judgment relative to his *own* evidence.
- But *advice* is also possible, in contrast to *subjectivism*. An advisor who knows more will assess the deliberator’s reason/ought judgment relative to *his* evidence, and so will disagree with it.
- At the same time, both deliberator and advisor will be trying to answer the *same* question, “What ought the deliberator do/believe?” Indeed, they can be disagreeing over the very same proposition, e.g., that the deliberator ought to leave both shafts open. So they will not talk past each other.

References: Kolodny and MacFarlane, “Ought: Between Objective and Subjective,” “Ought to Believe: Between Truth and Justification”

### **Implication: modus ponens is invalid**

Imagine that we don't know where the miners are. Then we might say:

A1 If the miners are in A, then it is not the case that we ought to leave both shafts open (since we ought, instead, to block A).

A2 We ought to leave both shafts open.

But we surely can't conclude:

A3 Therefore, the miners are not in shaft A, by modus tollens (and so modus ponens).

Again:

B1 If the miners are in A, then it is not the case that we ought to leave both shafts open.

B2 If the miners are in B, then it is not the case that we ought to leave both shafts open.

B3 Either the miners are in A, or the miners are in B.

But we surely can't conclude:

B4 Therefore, it is not the case that we ought to leave both shafts open, by constructive dilemma (and so modus ponens).

So these arguments are invalid. Why?

- “Because they equivocate. The ‘ought’ in A1 is ‘objective,’ whereas the ‘ought’ in A2 is ‘subjective.’” But we've already rejected the idea that ‘ought’ has objective and subjective senses. That can't make sense of disagreement and retraction.
- Because modus ponens is not valid for the indicative conditional.

When is “If P, then Q” true? Here's an approximation:

1. Think of  $i$ , the body of information, as the set of worlds compatible with what is known by someone with that information: all the ways the actual world might be, given that knowledge.  $i$  contains the actual world,  $w$ .
2. Remove from  $i$  all of the worlds at which  $P$  is false. This gives us a new set,  $i+P$ . Roughly,  $i+P$  is what would be known by someone who started with  $i$  and then learned  $P$ .
3. Then ask: Is  $Q$  true at  $\langle w', i+P \rangle$  for every  $w'$  in  $i+P$ ? Roughly, would that person, who started with  $i$  and then learned  $P$ , know  $Q$ ? (Compare Ramsey's test.)
4. If (and only if) the answer is yes, “If  $P$ , then  $Q$ ” is true.

*How modus ponens can lead us astray:*

This makes it possible for  $P$  and “If  $P$ , then  $Q$ ” to be true, but  $Q$  to be false: for modus ponens to be invalid. This happens when:

1.  $P$  is true and  $Q$  is false at  $\langle w, i \rangle$  but
2.  $Q$  is true at  $\langle w', i+P \rangle$  for all  $w'$  in  $i+P$ . (I.e., “If  $P$ , then  $Q$ ” is true.)

This happens with the conditional in our paradoxical arguments:

1.  $Q$  is false at  $\langle w, i \rangle$ . Given what little we know ( $=i$ ), we ought to leave both open.
2. Perfectly compatible with this that  $P$  is true at  $\langle w, i \rangle$ : that the miners are in A at the actual world,  $w$ .

3. Q is true at  $\langle w', i+P \rangle$  for all  $w'$  in  $i+P$ . If we knew that the miners were in A ( $=i+P$ ), then given what we knew, it would not be the case that we ought to leave both open. (Coming to know P turns Q from false to true.)

*Why modus ponens usually doesn't lead us astray:*

The possibility we've been considering *cannot* happen when either of the following conditions is met:

1. Q's truth depends only on the world, not on the body of information. Then if Q is true at  $\langle w, i+P \rangle$ , then Q had to be true at  $\langle w, i \rangle$  too. (Coming to know P can't turn Q from false to true, if Q's truth *doesn't depend* on what is known.)
  - C1 If the miners are in shaft A, they have a jackhammer.
  - C2 They do not have a jackhammer. (Whether this is true just depends on  $w$ , not  $i$ : on how the actual world is, not on what is known about it.)
  - C3 So, they are not in shaft A.
  
2. P is already known at  $i$ . Then  $i=i+P$ . Then if Q is true at  $\langle w, i+P \rangle$ , Q had to be true at  $\langle w, i \rangle$  too. (Coming to know P can't turn Q from false to true, if P is *already* known.)
  - D1 If it's raining, the streets must be wet.
  - D2 It's raining. (This is something known.)
  - D3 So, the streets must be wet.

One of these conditions is usually met. So modus ponens is usually safe to follow. This explains why modus ponens *seems* valid.

*How modus ponens can lead us astray, continued*

Modus ponens can lead us astray only when neither of the above conditions is met. That is, when:

1. Q's truth does depend on the body of information, and
2. P is not known at  $i$ .

This is what happens in the paradoxical arguments. Fully spelled out, the first argument is:

- E1 If the miners are in A, then it is not the case that we ought to leave both shafts open. (Here Q is an 'ought'-claim, which depends on what we know.)
- E2 We ought to leave both shafts open.
- E3 Suppose (for reductio) that the miners are in A. (Here we don't know P; we're only supposing it for the sake of argument.)
- E4 Then it is not the case that we ought to leave both shafts open. Contradiction. (Here we apply modus ponens where both conditions fail.)
- E5 So, by reductio, it's not the case that the miners are in shaft A.

Reference: Kolodny and MacFarlane, "Ifs and Oughts"

### **Implication: An argument that reasons are facts is questionable**

Now let us consider the claim that reasons are facts.

Again, this is *not* the same thing as objectivism: the claim that what there is *reason* for depends on *all* of the facts. A subjectivist, who holds that the subject's evidence is just what he *knows*, might agree that *reasons* are facts (since what the subject knows are facts).

One ground for thinking that reasons are facts is the weirdness of:

#“(i) There is a reason to believe P, namely Q. (ii) But Q is not the case.”

This weirdness would be explained by the hypothesis that reasons are facts.

One might try to explain the weirdness, instead, by claiming that (i) means:

“(i\*) I would be justified in believing P because I believe Q.”

and that's Moore's Paradox. That's why it's weird.

But as we've seen, (i) does *not* mean (i\*). A third party will assess (i) and (ii) as contradictory, but need not assess (i\*) and (ii) as contradictory.

However, there is another explanation of the weirdness. Suppose:

- a. “There is a reason to believe P, namely Q” is true relative to a body of information only if Q belongs to that information.
- b. An assessor's body of information is *everything that assessor believes, whether or not it is true*.

Suppose *you* say (i) and (ii). Relative to *my* information, as assessor, (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 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 facts. But that doesn't mean that they are.

## Rational Requirements

“Rational requirements”: Requirements that depend exclusively on *how things are with the subject*, requirements that the subject believe or intend *what coheres with his or her own attitudes or evidence*.

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)
- 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. *Subjective evidence requirements*: that we believe or intend in accord with what is likely *given our evidence*. These are just equivalent to the *subjective 'ought,'* to what it would be *reasonable* for one to do, or what one would be *justified* in believing.

- 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*—or 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.") This seems to me compatible with a variety of views about what evidence is: what is known, fixed by perceptual experiences, what is supported by the bulk of what one believes, etc.

### **Puzzle for rational requirements**

- (1) Rational requirements seem to guide deliberation and to offer advice. So it is natural to think that we have reason, and ought, to comply with rational requirements. Rational requirements are just a species of ought, supported by reasons. After all, reason and ought, as we are understanding them, are just what matter in deliberation and advice.
- (2) The following substantive views about what matters in deliberation and advice, about what we have reason or ought to do, seem plausible.

*Evidentialism*: There is reason for one to *believe* P to the extent that it is likely that P is true.<sup>1</sup> There is reason for one to *refrain from believing* P to the extent that it is likely that P is false. (Reasons to believe something show that it is likely to be true.)

*Value-based*: One has reason to E to the extent that one's E-ing would realize a valuable state of affairs (or avoid a disvaluable state of affairs) or respect something of value.<sup>2</sup> (Reasons for ends are that those ends realize valuable states of affairs, or respect things of value.)

*or*:

---

<sup>1</sup> Might add: "...and that P would still be true if one were to believe it and one will have, or there is sufficient reason for one to have, an opinion whether P."

<sup>2</sup> This may depend on our individual relations to things of value (e.g., my friendships may give me reasons that yours does not).

*Desire-based*: One has reason to E to the extent that one's E-ing would realize a state of affairs that one desires.

*General Transmission*: One has reason to M to the extent that one has reason to E and it is likely, if one M's, that one's M-ing helps to bring about one's E-ing in a nonsuperfluous way. (Reasons for means are that those means have some chance of helping to bring about some end for which one has reason.)

Reference: Kolodny, "Toward a Principle of Instrumental Transmission"

- (3) However, rational requirements tell us to do *otherwise* than what these substantive theories say: to do things other than to believe what is likely true, refrain from believing what is likely false, to do what is likely to realize or respect value, realize what we desire, etc. This happens in cases where:
- our evidence is misleading,
  - our normative beliefs are false,
  - we achieve formal coherence in a way that leaves us further from, rather than closer to, believing what is likely to be true, doing what is likely to promote or respect value, etc.

### **What I should have said**

In some earlier work, I drew this (admittedly artificial) distinction between reason and rationality in the following way.

- On the one hand, I said that reasons are *facts* and sometimes facts *other than that one has certain (nonfactive) attitudes*.
- On the other hand, I said that rational requirements are such that whether one satisfies them *depends only on one's (nonfactive) attitudes*, in abstraction from any other facts.

I now think that this was inapt, because:

1. This way of drawing the contrast suggests that what is important about *reason* is some metaphysical issue about *factivity*, rather than its functional role in *deliberation and advice*.
2. It isn't clear that reasons *are* facts, as we saw above.
3. This way of drawing the contrast suggests that what is important about *rationality* is its depending on how things *mentally* are with the subject, rather than how things are *with the subject*: mentally or in terms of evidence. It's the *subject-centeredness* of rationality, not the psychological-ness of rationality, that accounts for the puzzle: that makes it hard to reconcile rationality with reason, which *doesn't* depend (just) on how things are with the subject.
4. It isn't clear that rationality *does* depend exclusively on how things *mentally* are with the subject. This is because a subject's *evidence* might depend on more than simply his (nonfactive) attitudes (e.g., if all evidence is knowledge).

### **An error theory for subject's-evidence requirements**

1. *Confusion of the evaluative and the deliberative*: Subjective-evidence requirements may not be meant to play any role in deliberation and advice. Their role might instead

be merely as *evaluative* standards, which we use to *appraise* performances, as bystanders might. 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 she has responded to *her* evidence, not to ours.

2. From the *first-person* point of view, they *seem* like genuine *deliberative* requirements. When *I* assess *my own* claim “I ought to X,” its truth or falsity stands or falls with the truth or falsity of my claim “It would be reasonable for me to X.” When *I* am the assessor, there is no distinction between *my* evidence and *the* evidence.  
“You would be justified in believing P, given what you know. But you ought not believe P, because it can't possibly be the case.  
#“I would be justified in believing P, but I ought not believe P, because it can't possibly be the case.”

### An error theory for kratic requirements

#### 1. Confusing claims about patterns of reasons with claims about rational requirements

Here's a quick error theory for K+. The only reason K+ seems plausible to us, this error theory claims, is the:

*Violation Intuition*: If one believes that one ought to have A, but one does not have A, then one violates some requirement.

But to explain VI, we don't need:

K+: If one believes that one ought to have A, then one is required to have A.

We can instead appeal to the following:

- *Fact about reason patterns*: In any given case, either I ought not believe that I ought have A, or I ought to have A.

What explains *Fact*?

- *Epistemic Strictness*: If it is not the case that I ought to believe P, then I ought not to believe P.
- What I ought to believe and what I ought do depends on the same body of information. So if I ought to believe that I ought to have A, then I ought to have A.

But *Fact* doesn't explain the:

*Satisfaction Intuition*: Suppose one believes that one ought to have A, but does not have A. If one forms A, even though one ought not have A, then one satisfies a requirement, or gets something right, in a way that one would not have if one had not formed A.

By contrast, K+ would explain the Satisfaction Intuition.

#### 2. Some failed attempts to identify a reason to satisfy K+

“By satisfying K+ over the long run, one is more likely to conform to reason over the long run.”  
Even if this is true about the long run, in what sense do I do something I ought *in this particular case*?

“There just is some intrinsic value in satisfying K+.”

1. Implausible, as though psychic tidiness were an end in itself.
2. This intrinsic value would have to be practical: that it properly responds to some value or satisfies some desire. So as far as the required attitude, A, is a belief, this still wouldn't satisfy Evidentialism.

### 3. *Confusion of the evaluative and the deliberative:*

By satisfying K+, one functions well.

1. In what way is satisfying K+ *functioning well*, etc.? The answer may be that it is manifesting a disposition that is good to have, insofar as that disposition leads one to better conform to reason over the long run.
2. However, K+ *seems* to play a role in deliberation and advice that this doesn't account for.

### 4. *The Transparency Account*

1. K+ can *seem* to have significance for *first-person* deliberation: When K+ applies, the subject believes, of the response that K+ requires of her, that she ought to give it.
2. K+ can *seem* to have significance for *second-person* advice: The adviser can inform the subject of what is required by K+, with the expectation that, as a result of this advice, the subject will give the required response.

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

References: Kolodny, “Why Be Rational?” “Reply to Bridges”

## **A template for an error theory for formal coherence requirements, focused on N**

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 I ought to believe just what is sufficiently likely to be true (and not what is sufficiently likely to be false). How are these compatible? Suppose P is very likely to be true and not-P isn't. And suppose that I drop the belief that P, but keep the belief that not-P. Then I have satisfied N, but I'm further from what is likely true (and closer what is likely false).

Note that I might also be further from what *my* evidence suggests is true. This is just as much a problem for *subjectivism* about reason and ought. It has *nothing to do with objectivism* or my and MacFarlane's positive view that reason/ought/epistemic modals are assessment-sensitive.

### *1. Confusing claims about patterns of reasons with claims about rational requirements: confusing disjunctions of requirements with the requirement of a disjunction*

Again we might explain:

*Violation Intuition:* If one believes P and believes not-P, one violates some requirement.

Not with:

*N:* one is required (either not to believe P or not to believe not-P).

But instead with:

*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.

What explains *Fact*?

- *Stronger evidence*: One ought to believe P only if P is more likely than not-P.
- Possibly, *SE* is in turn explained by the fact that avoiding false belief is more important than acquiring true belief.

But, against, this fact wouldn'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 would explain the Satisfaction Intuition.

2. *Some failed attempts to identify a reason to satisfy N*

"By satisfying N over the long run, one is more likely to satisfy Evidentialism over the long run."

1. Again, 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 N, one does *not* better satisfy Evidentialism over the long run. (Complicated.) Formal coherence requirements are worse off than kratic requirements in this respect.

"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, this still wouldn't satisfy Evidentialism.

"In your example, 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 by dropping the belief that P, one does something that one ought to do. SI explained."

1. First, false that I ought to satisfy any necessary condition of what I ought to do. (Argued above.)
2. Second, another necessary condition of not believing not-P is (either not believing not-P or both believing P and believing not-P). By both believing P and believing not-P, one does something I ought to do. After all, anything satisfies a necessary condition of anything. But the Satisfaction Intuition was precisely that one satisfies a requirement *in a way that one would not have if one had continued believing P and believing not P*.

3. *Confusion of the evaluative and the deliberative*:

1. In what way is satisfying N *functioning well*, etc.? As we just saw, we can't answer: "Because it is manifesting a disposition that is good to have, insofar as that disposition leads one to better satisfy Evidentialism over the long run."
2. Formal coherence requirements *seem* to play a role in deliberation and advice that this doesn't account for.

#### 4. 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, then, given *Fact*, one must have some attitudes at odds with reason. So one ought to arrive at some belief about which of the relevant attitudes 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 (i.e., against reason), one still satisfies:

- i. Self-monitoring, which is a requirement of reason.
- ii. K+ and/or K-, whose seeming deliberative/advisory role we explained earlier.

Perhaps this explains the Satisfaction Intuition.

Objection: 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.

References: Kolodny, "How Does Coherence Matter?" "Why be Disposed to Be Coherent?"

### **Extending the Error Theory to other formal coherence requirements**

The discussion of N provides a template for error theories for other formal coherence requirements. What changes in each case is the "fact about reasons patterns" and the explanation of that fact. **Here are sketches of the explanations of C, IC, and ME.**

#### 1. C:

**Relevant fact about reasons patterns: When P entails Q, either I ought not believe P, or I ought to believe Q.**

**Explanation:**

- Suppose it is *not* the case that I ought not believe P. Then by...
- *Epistemic Strictness*
- ... I ought to believe P.
- *Evidential Transmission*: If Q must be the case if P, then Q is at least as likely as P.<sup>3</sup>
- Hence, if I ought to believe P, then I ought to believe Q.

**References: Kolodny, "How Does Coherence Matter?" "Why be Disposed to Be Coherent?"**

---

<sup>3</sup> I'm suppressing some complications here.

2. IC:

Relevant fact about reasons patterns: *Either* I ought not to believe that if I X, I do not Y, *or* I have *less* reason (to intend to X and to intend to Y) than I have for *some alternative* pattern of intentions, e.g., (to intend to X and not to intend to Y).

Example: I ought not to believe that if I go to London, I do not lecture in Berkeley, *or* I have *less* reason (to intend to go to London and to intend to lecture in Berkeley) than I have for *some alternative* pattern of intentions, e.g., (to intend to go to London and not to intend to lecture in Berkeley).

Explanation:

- Why do I speak of reasons for *patterns* of intentions?
  - *Reasons for intentions are, in general, instrumental.* That is, one's reason to intend to E is typically that intending to E is a means to E-ing.
  - According to *General Transmission*, one's reason for a means depends on how likely it is to bring about the end.
  - Various conditions may affect how likely it is to bring about the end. For example, how likely my intention to reach camp by nightfall is to bring it about that I do may depend on the weather. Thus, they affect the reason that one has to intend the end.
  - Among these conditions are *other* things that one is likely to do and their effects.
  - How likely one is to do these other things may depend on one's *other* intentions. This is what Scanlon calls the *predictive significance of intention*.
  - Since the reason that one has to intend to X may depend on whether or not one intends to Y, and vice-versa, it makes more sense to consider reasons for *patterns* of intention, e.g., (intending to X and intending to Y), (intending to X and not intending to Y), etc.
- Suppose it is not the case that I ought not believe if I go to London, I do not lecture in Berkeley. By Epistemic Strictness, I ought to believe that if I go to London, I do not lecture in Berkeley.
- Hence, it is likely that, if I go to London, I do not lecture in Berkeley.
- Suppose I intend to go to London
- If I intend to go to London, then I am likely to go to London.
- If I am likely to go to London, then I am not likely to lecture in Berkeley.
- If intending to lecture in Berkeley is unlikely to bring it about that I do lecture in Berkeley, then, by General Transmission, I have little reason to intend to lecture in Berkeley.
- In sum, if I intend to go to London, then I have little reason to intend to lecture in Berkeley.
- Presumably, there is reason against intending to lecture in Berkeley. (For example, it might lead me to take pointless means, such as preparing my lecture.)
- As a result, I have less reason (to intend to go to London and to intend to lecture in Berkeley) than I have, for example, (to intend to go to London and not to intend to lecture in Berkeley).

Reference: Kolodny, "The Myth of Practical Consistency"

### 3. ME?

Relevant fact about reasons patterns: *Either* I ought not to believe (I E only if I intend to M), *or* I have less reason (to intend to E and not to intend to M) than for some other pattern of intentions.

Example: *Either* I ought not to believe (I go to London only if I intend to buy a ticket), *or* I have less reason (to intend to go to London and not to intend buy a ticket) than for some other pattern of intentions.

Explanation:

- Suppose it is not the case that I ought not believe (I go to London only if I intend to buy a ticket). By Epistemic Strictness, I ought to believe it.
- Hence, it is likely that, I go to London only if I intend to buy a ticket.
- Suppose I do not intend to buy a ticket.
- This is, in effect, to suppose that it is unlikely that I intend to buy a ticket.
- Hence, it is very unlikely that I go to London.
- Hence, by General Transmission, there is little reason to intend to go to London, since it is unlikely to bring it about that I do go.
- In sum, if I do not intend to buy a ticket, then I have little reason to intend to go to London.
- Presumably, there is some reason against intending London.
- As a result, I have less reason (not to intend to buy a ticket and to intend to go to London) than I have (not to intend to buy a ticket and not to intend to go to London).
- This does *not* mean that I ought (not to intend to buy a ticket and not to intend to go to London) all things considered. For example, it might be the case that I ought (to intend buy a ticket and to intend go to London). So long as I *do* intend to buy a ticket, there may well be reason to intend to go to London. But it does mean that whatever pattern of intention I ought to have, it is *not* (not to intend to buy a ticket and to intend to go to London). There is a superior alternative.

References: Kolodny, "How Does Coherence Matter?" "Why be Disposed to Be Coherent?"