Notes on Chapters 6 & 7 of Schroeder's *Noncognitivism in Ethics*
("The Frege–Geach Problem, 1973–2006")

Branden Fitelson

September 15, 2016

1   Preamble: Some Notation, Terminology, and Dialectical Setup

In the interest of space, I'll be skipping over Blackburn's (earlier) "involvement" account, and jumping straight into his HOA account (§6.5 of Schroeder). But, first, I'll begin by introducing some notation and terminology that I'll be using throughout (note: I'll be following Schroeder rather closely again here).

$[p] \stackrel{\text{def}}{=}$ the mental state (or propositional attitude) *thinking that p*.

$\left\| m \right\| \stackrel{\text{def}}{=}$ the set of possible worlds in which mental state m is *correct/accurate*.

I will illustrate the main uses of this notation (and one other piece of notation to be introduced later) by discussing cognitivist (*viz.*, realist) *vs.* noncognitivist explanations of the following two central *explananda*.

The Inconsistency Property (INC). Some combinations of attitudes are *rationally inconsistent* (note: I'll understand "rational inconsistency" in terms of *epistemic incoherence/irrationality*).

The Inference-Licensing Property (ILP). Some combinations of attitudes *commit* rational agents (perhaps, *implicitly*) to other attitudes (*i.e.*, "inference-licensing" *via rational epistemic commitment*).

For instance, let P be some simple moral claim (*e.g.*, "murder is wrong") and let Q be some simple nonmoral claim (*e.g.*, "it's raining"). And, consider the *disjunction introduction* argument $P \therefore P \vee Q$. The realist will explain the validity of this argument in the usual, truth-conditional way. And, this will undergird the following cognitivist explanations of the corresponding instances of the two explananda (INC) and (ILP).

INC Explanandum: the combination of attitudes $\{[P], [\neg(P \vee Q)]\}$ is *rationally inconsistent* (*viz.*, *epistemically incoherent/irrational*).

ILP Explanandum: the attitude $[P]$ *commits* a rational agent (perhaps *implicitly*) to the attitude $[P \vee Q]$.

The (traditional) cognitivist explanation of this instance of (INC) goes as follows. Consider the set of possible worlds in which *both* of the attitudes $\{[P], [\neg(P \vee Q)]\}$ are *correct/accurate*. This is the set:

$$\left\| [P] \right\| \cap \left\| [\neg(P \vee Q)] \right\| = \varnothing$$

That is, there are *no* possible worlds in which *both* of the attitudes $[P]$ and $[\neg(P \vee Q)]$ are correct/accurate. Moreover, traditional cognitivists accept the following principle regarding epistemic rationality.[1]

Inconsistency Implies Incoherence (III). It is *epistemically irrational* to adopt a set of attitudes that (one knows) cannot all be correct/accurate/true (*i.e.*, a logically *inconsistent* set of attitudes).

(III) implies that the combination of attitudes $\{[P], [\neg(P \vee Q)]\}$ is *epistemically incoherent/irrational*.   □

The challenge for noncognitivists is to offer a general strategy for explaining (instances of) INC. Preferably, this strategy should have certain explanatory virtues (exhibited by the cognitivist explanatory strategy).[2] Next, I will go through some of the *expressivist* INC explanatory strategies Schroeder discusses in chapters 6 and 7. In the Epilogue, I will call into question some of the background epistemology (*e.g.*, assumption III above) that is presupposed in this dialectic. That will lead to a different way of thinking about INC and how to explain it (from both cognitivist and noncognitivist perspectives). This will also anticipate some of the epistemological issues Sarah Moss will be discussing when she visits the seminar in November.

---

[1]I will call this principle into question later. But, I will assume it for most of the handout.

[2]I think INC explananda are more (epistemologically) fundamental than ILC explananda. And, I think "implicit commitment" is epistemically murkier than "incoherence." So, I will (in the interest of space) focus exclusively on INC explanations in this handout.

2 Blackburn's "Higher Order Attitudes" (HOA) Semantics

Blackburn offers a higher-order attitude (HOA) expressivist account of moral thought (and talk). Specifically, he offers a *constructive recipe* for determining complex attitudes from their simple constituents. Let's start with a simple moral claim: ϕing is wrong. Blackburn proposes the following (atomic) recipe.

$$[\phi\text{ing is wrong}] := \text{DIS}(\phi\text{ing})$$

In words, this says that the attitude one has toward the proposition that ϕing is wrong is the attitude: DISapproval of ϕing, where it is understood that DISapproval has a *world-to-mind* direction of fit. Next, Blackburn proposes a general strategy for explaining INC. I will illustrate the strategy by looking at an example (but the strategy can be generalized). Consider the following *conditional* statement.

$$\phi\text{ing is wrong} \rightarrow \psi\text{ing is wrong}.$$

Truth functionally, this conditional is equivalent to the following statement.

$$\neg(\phi\text{ing is wrong \& } \neg(\psi\text{ing is wrong}))$$

Blackburn proposes a *higher-order* explication (which *sort of* mimics this structure). Let $\Phi\overline{\Psi}$-ing be the "mental act" of thinking that ϕing is wrong *while* (also) thinking that *it is not the case that* ψing is wrong.

$$\text{DIS}(\Phi\overline{\Psi}\text{-ing})$$

In words, the attitude one takes toward the proposition ⌜if ϕing is wrong *then* ψing is wrong⌝ (when one *thinks* it) is the *second-order* attitude of DISapproval of *having* the following *pair* of (first-order) attitudes: $\{[\phi\text{ing is wrong}], [\neg(\psi\text{ing is wrong})]\}$. More generally, Blackburn's recipe for conditionals can be stated as:

$$[p \rightarrow q] := \text{DIS}(p\overline{q}\text{-ing}) = \text{DIS}(\text{thinking that } p \text{ *while* thinking that } \neg q).$$

The good news is that this allows Blackburn to provide a *general, constructive* explanation of (many) INC explananda involving conditional statements. For instance, consider the following (traditional) INC explanandum: the combination of attitudes $\{[P], [P \rightarrow Q], [\neg Q]\}$ is *inconsistent*. On Blackburn's account, this combination of attitudes becomes: $\{[P], [\neg Q], \text{DIS}(P\overline{Q}\text{-ing})\}$. And, plausibly, it is *rationally incoherent* to (simultaneously) have both of the first-order attitudes $[P]$ and $[\neg Q]$, *while having a second order attitude of disapproval toward* simultaneously having both of these first-order attitudes (think: *epistemic akrasia*).

The bad news is (at least) three-fold. First, there is no adequate way to *generalize* this (single DIS attitude) account, because of **the negation problem** (I'll return to the negation problem in section three below). Second, Blackburn's account yields *incorrect analyses* of *purely descriptive* conditionals. To see this, consider

(DD) If the Bible tells us to be unfriendly, then my parents lied to me.

On Blackburn's account, my thinking that (DD) *just is* my DISapproving of being in a certain state of mind (*viz.*, the state of mind of believing that the bible says not to be friendly while denying that my parents lied to me). But, since the antecedent and consequent of this conditional are both nonmoral (*i.e.*, descriptive), this just seems to *get the meaning of the conditional wrong*. Intuitively, the (truth functional) meaning of this conditional is: *either* the bible does *not* tell us to be unfriendly *or* my parents lied to me.

The third problem for Blackburn's HOA account is that (from a traditional INC perspective) it *over-generates* "validities" (and, hence, it also *over-generates* "inconsistencies"). Let $\Phi \stackrel{\text{def}}{=} \phi\text{ing is wrong}$, and $\Psi \stackrel{\text{def}}{=} \psi\text{ing is}$ wrong. As van Roojen explains, Blackburn's account implies that the following two claims are *equivalent*.

(*A*) If Φ, then Ψ.

(*B*) It is irrational/incoherent to think that Φ *while* thinking that $\neg\Psi$.

Intuitively, the argument $\Phi, A \therefore \Psi$ is *valid*, and so the set $\{\Phi, A, \neg\Psi\}$ is *inconsistent*. If Blackburn's account were correct, then the argument $\Phi, B \therefore \Psi$ would (also) be valid, and so the set $\{\Phi, B, \neg\Psi\}$ would (also) be inconsistent. But, intuitively, neither of these claims is correct. So, Blackburn's account over-generates "validities" and "inconsistencies," which leads to *too many INC explananda*. Schroeder diagnoses van Roojen's problem as (partly) a problem of specifying the "right kind" of rational inconsistency/incoherence. He says

> ...if expressivists are going to explain a version of [the INC] that suffices to distinguish valid arguments from invalid arguments, then they can't appeal to just any old kind of rational conflict between mental states — they need to appeal to the kind of clash that obtains between beliefs with inconsistent contents. ...we've already put some thought ...into whether desire-like attitudes can clash with each other in this way. It was what we called "intrapersonal disagreement."

As I will explain in the Epilogue, I think this is misguided (because it misunderstands the nature of epistemic incoherence and its relation to logical inconsistency). But, for the purposes of my discussion of Gibbardish semantics in the next section, I will assume there is a suitable notion of "intrapersonal disagreement" which plays the role (*viz.*, a notion which is suitably analogous to logical inconsistency) that Schroeder wants.

3 Gibbardish Semantics

In order to grok Gibbardish semantics, we'll need to introduce a new piece of notation.

$|m| \overset{\text{def}}{=}$ the set of mental states that you land yourself in disagreement with by being in mental state m.

For instance, $|[p]|$ is the set of mental states you land yourself in disagreement with *by thinking that p*. Intuitively, $|[p]|$ *includes* $[\neg p]$. That is, intuitively, $[\neg p] \in |[p]|$. But, that doesn't yet *characterize* $[\neg p]$, as a function of $[p]$. We'll return to negation below. Let's start with disjunction. What is $|[p \vee q]|$, as a function of $|[p]|$ and $|[q]|$? Intuitively, $[p \vee q]$ lands one in *fewer* disagreements than either $[p]$ *alone* or $[q]$ *alone* does. On the traditional (inconsistency-based) account of "disagreement," we have the following.

$$|[p \vee q]| = |[p]| \cap |[q]|$$

In words, this says that those states which land one in disagreement with *both $[p]$ and $[q]$* are precisely the states which land one in disagreement with $[p \vee q]$. This is, basically, a *De Morgan Law* for "disagreement." We can mimic other truth functional laws as well, in precisely the ways one would expect. In order to complete the (Boolean) analogy, we need only characterize *negation*. If $|[p]| \subseteq |[q]|$, then $[q]$ is "at least a strong as" $[p]$, (in the Boolean-analogical sense of "disagreement"). That is, if $q \vDash p$, then $[q]$'s *disagreement class* is a *superset* of $[p]$'s (in the intended sense). With this in mind, we can characterize negation as follows.

$$|[\neg p]| = \left\{ y \mid |[p]| \subseteq |y| \right\}$$

Once we have clauses for negation and disjunction, we can then give the Gibbardish semantics of all other (Boolean-analogue) complexes in the usual way. So, for instance, the conditional would be given by.

$$|[p \to q]| = |[\neg p \vee q]| = |[\neg p]| \cap |[q]| = \left\{ y \mid |[p]| \subseteq |y| \right\} \cap |[q]|$$

It is straightforward to show that these semantic clauses endow "disagreement" with a "logic" (*i.e.*, a formal "disagreement structure") that mirrors classical/Boolean logic (*i.e.*, classical Boolean inconsistency). In other words, for every classical logical validity (and its corresponding inconsistency), there exists a (unique) corresponding "validity"/"inconsistency" in terms of "disagreement classes." Because Gibbardish semantics satisfies this *formal* constraint, it avoids van Roojen's "over-generation of validities/inconsistencies" problem. However, the account still faces some serious challenges (both formally and philosophically).

First, as it stands, the account is *non-constructive*. It is non-constructive because (unlike the HOA account) it doesn't tell us *which* mental states (*if any*) satisfy the right hand sides of the semantic clauses. All we have been given are are *descriptions* of mental states, and what their "disagreement classes" *would have to be like*, *if* such states exist and *if* Gibbardish semantics is going to explain (all) the INC explananda in a way that is (formally) analogous to the classical logical explanation. Here, expressivism faces a *trade-off* between *formal* adequacy and *constructivity*. Indeed, this trade-off is *unavoidable*, because of **the negation problem**.

The negation problem can be explained as follows. Intuitively, "ϕing is wrong" and "ϕing is not wrong" (*i.e.*, "it is not the case that ϕing is wrong") are inconsistent, as are "not-ϕing is wrong" and "not-ϕing is not wrong." But, "ϕing is not wrong" and "not-ϕing is not wrong" are *not* inconsistent (since ϕing and

not-ϕing may both be permissible). However, it can be shown that any (constructive, single DIS attitude) approach which explains the former (*true*) inconsistencies will imply this latter (*false*) inconsistency (Schroeder, pp. 135–36). So, in order to salvage formal adequacy, one will need to postulate (infinitely) *many* "DIS attitudes," which renders the account highly non-constructive (and also comparatively *complex* and *ad hoc*).

Because Gibbardish semantics are so non-constructive, there is an important sense in which they fail to be *as explanatory* as the classical truth functional/cognitivist approach (or even the HOA approach). The classical explanation of the incoherence of the pair of attitudes $\{[p],[\neg p]\}$ tells us precisely what the contents of these two attitudes are, and *why* they are incoherent. Gibbardish accounts simply *assume* that, *given* a mental state $[p]$, *there exists* some mental state that "disagrees" with it (in the requisite, "intrapersonal" sense), which they call "$[\neg p]$." But, such accounts don't tell us *what* $[\neg p]$ *is*, or *why* it "disagrees" with $[p]$.

4 Epilogue: Inconsistency and Incoherence/Irrationality Revisited

There are other problems that arise from *epistemological background/auxiliary assumptions* about the relationship between logical constraints and rational requirements. We have already seen one of these.

> **Inconsistency Implies Incoherence/Irrationality** (III). It is *epistemically irrational* to adopt a set of attitudes that cannot all be correct/accurate/true (*i.e.*, a logically *inconsistent* set of attitudes).

I think (III) is false. And, I'm not alone. Foley, Christensen, and others also think (III) is false. They have a different understanding of what epistemic rationality requires; one which is compatible with there being *inconsistent* sets of beliefs that are *rationally jointly held*. My favorite example is the following twist on the (traditional) *preface paradox*, which is discussed in a paper that Kenny Easwaran and I published last year.

> **Revenge Preface**. John is an excellent empirical scientist. He has devoted his entire (long and esteemed) scientific career to gathering and assessing the evidence that is relevant to the following first-order, empirical hypothesis: (*H*) all scientific/empirical books of sufficient complexity contain at least one false claim. By the end of his career, John is ready to publish his masterpiece, which is an exhaustive, encyclopedic, 15-volume (scientific/empirical) book which aims to summarize (all) the evidence that contemporary empirical science takes to be relevant to *H*. John sits down to write the Preface to his masterpiece. Rather than reflecting on his own fallibility, John simply reflects on the contents of (the main text of) his book, which constitutes *very strong inductive evidence in favor of H*. On this basis, John (inductively) infers *H*. But, John also believes each of the individual claims asserted in the main text of the book. Thus, because John believes (indeed, knows) that his masterpiece instantiates the antecedent of *H*, the (total) set of John's (rational/justified) beliefs is inconsistent.

So, Inconsistent (set of belief contents) $\nRightarrow$ (Epistemically) Irrational (set of beliefs). As such, it seems like a mistake to "reverse engineer" our notion of "intrapersonal disagreement"/"rational inconsistency" so as to ensure that it lines up with classical *logical inconsistency*. But, this is precisely the "formal adequacy requirement" that seems to be presupposed in this dialectic. I like the move on the part of the non-cognitivists to focus on *epistemic irrationality*, rather than *logical inconsistency per se*. But, I would encourage an *even more radical move* away from anything that has even the *formal* structure of classical logical inconsistency, since I don't think *that* is the formal structure of epistemic incoherence (or "rational inconsistency").

Instead, I would propose a different formal requirement. Specifically, Kenny and I propose the following necessary requirement of epistemic rationality (at least, for a large class of belief content types).

> **Weak Lockean Coherence** (WLC). Any set of belief contents held by a rational agent *S* should be *Lockean-representable* (with a threshold of $1/2$) by *some* probability function. In other words, there should exist some probability function $\Pr(\cdot)$ such that, for each p believed by S, $\Pr(p) > 1/2$.

(WLC) is a *formal* requirement, but it's *strictly weaker than logical consistency*. If we adopt (WLC) as our formal epistemic rational requirement, then this will significantly *shrink* the set of (INC) *explananda*. Some of the existing explananda will remain (*e.g.*, incoherence of $\{[p],[\neg p]\}$). But, many others will *not*. Also, the *explanans* of our explanations of epistemic incoherence will become *more complex* (and *probabilistic*!). This anticipates some of the moves Sarah Moss will be making in her new book (parts of which we'll read soon).