## Contents

<table>
<thead>
<tr>
<th>1</th>
<th>Contractarian Ethics</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>The Source of Morality</td>
<td>5</td>
</tr>
<tr>
<td>1.1.1</td>
<td>Religious</td>
<td>5</td>
</tr>
<tr>
<td>1.1.2</td>
<td>Self-Interest</td>
<td>6</td>
</tr>
<tr>
<td>1.1.3</td>
<td>Ethology</td>
<td>6</td>
</tr>
<tr>
<td>1.1.4</td>
<td>Contractarianism</td>
<td>7</td>
</tr>
<tr>
<td>1.2</td>
<td>The State of Nature</td>
<td>8</td>
</tr>
<tr>
<td>1.2.1</td>
<td>Thomas Hobbes</td>
<td>8</td>
</tr>
<tr>
<td>1.2.2</td>
<td>John Locke</td>
<td>11</td>
</tr>
<tr>
<td>1.2.3</td>
<td>Jean-Jacques Rousseau</td>
<td>12</td>
</tr>
<tr>
<td>1.3</td>
<td>Modern Contractarianism</td>
<td>14</td>
</tr>
<tr>
<td>1.3.1</td>
<td>John Rawls</td>
<td>15</td>
</tr>
<tr>
<td>1.3.2</td>
<td>David Gauthier</td>
<td>16</td>
</tr>
<tr>
<td>1.4</td>
<td>Critiquing Contractarianism</td>
<td>18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2</th>
<th>Consequentialist Ethics</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Consequentialism, In General</td>
<td>21</td>
</tr>
<tr>
<td>2.2</td>
<td>Consequentialism, In Particular: Utilitarianism</td>
<td>23</td>
</tr>
<tr>
<td>2.3</td>
<td>An Objection to Utilitarianism</td>
<td>24</td>
</tr>
<tr>
<td>2.4</td>
<td>An Objection to Consequentialism</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3</th>
<th>Deontological Ethics</th>
<th>27</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Overview</td>
<td>27</td>
</tr>
<tr>
<td>3.2</td>
<td>Kantian Ethics</td>
<td>27</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Reason, Freedom, and The Good Will</td>
<td>28</td>
</tr>
<tr>
<td>3.2.2</td>
<td>‘The Formula of Universal Law’</td>
<td>29</td>
</tr>
<tr>
<td>3.2.3</td>
<td>‘The Formula of the End in Itself’</td>
<td>30</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Empty Formalism?</td>
<td>31</td>
</tr>
<tr>
<td>3.3</td>
<td>Agent-Relativity and Agent-Neutrality of Value</td>
<td>32</td>
</tr>
<tr>
<td>3.3.1</td>
<td>John Rawls</td>
<td>32</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Bernard Williams</td>
<td>33</td>
</tr>
</tbody>
</table>
3.4 Prima Facie Duties ................................................. 36

4 Metaethics: A Quick Introduction .......................................................... 41
  4.1 Taxonomy ............................................................... 41
  4.2 Cognitivism versus Non-cognitivism ........................................... 43
    4.2.1 Why Be Non-cognitivist? ........................................ 43
    4.2.2 Why Worry about Non-cognitivism? .......................... 49
  4.3 Naturalism versus Non-naturalism .................................................. 50
    4.3.1 Why be Non-naturalist? .......................................... 51
    4.3.2 Why Worry about Non-naturalism? ............................ 53

5 Relativism, Objectivism, and Humeanism ................................................. 55
  5.1 Relativism ............................................................. 55
    5.1.1 Descriptive Moral Relativism .................................... 56
    5.1.2 Functional Necessity ............................................. 57
    5.1.3 Persistent Disagreement ........................................ 57
    5.1.4 Fundamental Diversity .......................................... 58
    5.1.5 Internalism ....................................................... 59
  5.2 Objectivism .......................................................... 60
    5.2.1 Relativism Is Contradictory .................................... 60
    5.2.2 Relativism Makes Moral Disagreement Impossible .......... 61
    5.2.3 Relativism Is No Better than Objectivism .................. 61
  5.3 Humeanism ............................................................ 62
    5.3.1 Motivation ....................................................... 63
    5.3.2 Rationality and Reasons ....................................... 63
Chapter 1

Contractarian Ethics

1.1 The Source of Morality

Suppose we’re interested in morality’s source. This could mean that we are interested in the answer to either of these questions:

1. Why do we have the moral rules (“norms”) that we have? (A scientific question, answerable by biology, psychology, maybe game theory.)

2. What, if anything, justifies or grounds the norms that we have? Why should we accept these norms? Put simply: why be moral? (A non-scientific, normative question.)

When I say we’re interested in the source of morality, I mean that we’re interested mainly in finding an answer to the second of these questions. We can call this the Task of Justifying Morality.

Let’s quickly overview some venerable accounts of morality’s source (We’ll breeze through a few of these and come back to them in more detail later on. Note: these categories can overlap):

1.1.1 Religious

According to religious accounts, a moral rule R is justified by the fact that God has established it. But what reason is there to care about what God has established? If there is some independent reason, then the fact that God established R isn’t ultimately what grounds it. If there’s not, then the fact that God established R should render the question why be moral? moot. But it doesn’t.


CHAPTER 1. CONTRACTARIAN ETHICS

1.1.2 Self-Interest

According to these accounts, being moral makes your life go better than it otherwise would (Socrates, Plato, Hobbes). The need to appeal to a self-Interest account is often supported with PSYCHOLOGICAL EGOISM (the claim that people are, in fact, motivated only by their own interests) and the idea that justifying a moral rule involves supplying a motive (reason) to follow it.

Some worries about them:

- This seems to assume that to justify a moral rule to someone, we need to show them how it benefits them. As Midgley suggests (1993, p. 5) this seems too strong—we can (and often do) appeal to other motives besides self-interest (loyalty, compassion, pity, etc.) in order to justify moral rules. (How might the egoist respond to this?)

- The factual claim seems wrong—being moral often involves acts of altruism.

1.1.3 Ethology

Something like this is MIDGLEY’s view. According to her, our moral rules are the expression of behavioral dispositions (especially the desire for social interaction) that (i) we share with other higher mammals, (ii) have been naturally selected because they promote survival of our genes. These dispositions “supply the raw material of the moral life—the general motivations which lead towards it and give it its rough direction” (p. 9).

While the evolutionary origins of morality are profoundly interesting—and are probably the only way of explaining why we have the moral rules we have (and making scientific sense of our practice of evaluating our actions and the actions of others using a moral vocabulary)—it’s not clear how they supply a justification for morality. What reason is there to care about the moral rules that evolution underwrites? (We have precisely the same problem that afflicted the religious justification.)

We’ll consider evolutionary accounts in more detail at the end of the course.

---

1 As Midgley notes (p. 8), the evolutionary biological theory that genes are selfish (see Richard Dawkins, The Selfish Gene) should not be confused with the idea that selfishness (indifference to the interests of others) promotes survival of our genes. Qualities like altruism and trustworthiness (and moral rules that recommend them) are actually survival-promoting, as we’ll see when we talk about the Prisoner’s Dilemma.
1.1.4 Contractarianism.

A natural idea—which occurred to many famous philosophers (THOMAS HOBBES, JOHN LOCKE, JEAN-JACQUES Rousseau among them) and which we’ll spend a lot of time on—is that what justifies a moral rule is the fact that it has been agreed to by people that are similar, in relevant respects, to us (or, better, would be agreed to by people that have qualities that we judge to be important in deciding, as a group, what the moral rules are going to be). We’ll call this kind of theory contractarian, since the actual or hypothetical existence of an agreement (contract) is what they take to be crucial to justifying moral rules. The people whose agreement we’re interested in will be called the contractors.

An important way that contractarian accounts differ from one another is in how they answer the following question: what are the qualities that it is important for the contractors to have? Some possible answers (some of which may be combined):

- Self-interested (they care about nothing but their own wellbeing)
- Impartial (they care equally about everyone)
- Ignorant (they don’t know certain facts about themselves—e.g., gender, class)

Discussion: why is a contractarian account’s answer to this question so important? (Remember: we’re interested in justifying moral rules. The fact that others have placed themselves under an agreement doesn’t mean we’re bound by that agreement. So we need some other reason to think we should care about the agreement that these people have reached. But why care that a random group of yahoos reaches an agreement that some moral rule is a good rule for them to adopt, unless that random group of yahoos is designed to be sort of group whose decision we have reason to take seriously?)

Note: a contractarian account of morality’s source may not tell us what the moral rules actually are (or even how we might go about determining what they are). There are two ways for this phenomenon (MORAL INDETERMINACY) to occur.

1. The qualities that it is important for the contractors to have might be (although they don’t have to be) compatible with several different agreements. A description of the bargaining situation might not, all by itself, determine what agreement is reached. (Think of a contract that sets up a system of government. One group of self-interested contractors might
CHAPTER 1. CONTRACTARIAN ETHICS

decide on an absolutist monarchy, if they’re security-conscious, while another group of self-interested contractors might decide on a democracy, if they’re equality-conscious.)

2. The agreements themselves might be compatible with different moral codes. (Think again of a contract that establishes a system of government—e.g., representative democracy. The agreement is definite, but it doesn’t necessarily determine what the laws actually are. It might establish a procedure for deciding what the laws actually are, and stipulate that any moral code that is the outcome of such a procedure is just.)

This depends on what qualities we insist on for the contractors. If we’re specific about the qualities we insist on, we might avoid both kinds of indeterminacy.

1.2 The State of Nature

The first group of contractarian theories we’ll focus on are what we might usefully call STATE OF NATURE theories. The State of Nature is the pre-contract status quo: it is, therefore, a state in which there are no binding moral rules (Hobbes, perhaps Rousseau), or only relatively few binding moral rules—the ones that are in place naturally (Locke). The SOCIAL CONTRACT is some sort of agreement among people in the State of Nature to establish a code of rules and thereby exit the State of Nature.

Most of the early social contract theorists use the Social Contract as a tool of political justification—a means of justifying civil society and the state. In other words, they use the Social Contract as a way of grounding political morality. We’re interested in the Social Contract as a more general tool (although thinking of the agreement as a political arrangement can be a useful way of making the content of the agreement a bit less abstract).

1.2.1 Thomas Hobbes

Hobbes (1588–1679) says the State of Nature is a state of “war of every man against every man” (by which Hobbes means, not a state of constant violence, but a state in which people are generally disposed to violence) where life is “nasty, brutish, and short” (aside from the constant fear of violence, there is no common enterprise of any sort—no infrastructure, only subsistence agriculture etc.).
1.2. THE STATE OF NATURE

Why does Hobbes think the State of Nature is like this?

"Equality of ability" (+PSYCHOLOGICAL EGOISM) ⇒
“Equality of hope” in attaining our goals ⇒
Competition for scarce resources ⇒
Invasion for gain (“diffidence”) ⇒
Desire to “augment dominion” beyond basic needs ⇒
Invasion for safety

So there’s a lot of violence and people are fearful.

Why no common enterprise? Consider the following sort of case, known as the PRISONER’S DILEMMA (such cases are often called “market failures”).

<table>
<thead>
<tr>
<th></th>
<th>A cooperates</th>
<th>A defects</th>
</tr>
</thead>
<tbody>
<tr>
<td>B cooperates</td>
<td>+7, +7</td>
<td>−10, +10</td>
</tr>
<tr>
<td>B defects</td>
<td>+10, −10</td>
<td>−7, −7</td>
</tr>
</tbody>
</table>

Discussion: what should B do: cooperate or defect? [Clearly defect. Suppose A cooperates. Then B should defect. Suppose A defects. Then B should defect. So no matter what, B should defect. What should A do? By the same reasoning, A should defect. So they both should defect. Assuming they’re both rational, they both will defect, leading to the State of Nature (−7, −7) box. Common Enterprise (the box where they both cooperate) is better for everyone, but it’s impossible to rationally choose it in the State of Nature.]

Suppose we repeat this case (“iterate” it). The players will find themselves in a NASH EQUILIBRIUM:2 A has her strategy (defect), B has hers (defect), and neither has any incentive to change her strategy unilaterally (i.e., without the other changing her strategy). Common Enterprise will never arise, unless we break the Equilibrium—common enterprise depends on reciprocation, and reciprocation is irrational. This requires altering the incentives (e.g., punishing for defecting).

What role is the SoN playing in Hobbes’ account? It seems to be doing all of the following:

1. Furnishing a normative contrast. It is a state where there are no binding moral rules. (As such, there is a reason to move to a state with binding moral rules.)

2Named after mathematician John Nash, of A Beautiful Mind fame.
2. Furnishing an explanation—an account of the etiology—of our moral rules.

3. Furnishing a description of the contractors.

Discussion: what are Hobbes’ contractors like? What kind of society do you think they would agree to?

In a pre-contract state (one without binding moral rules), people are narrowly self-interested (see the Prisoner’s Dilemma), fearful, concerned to get out of that state, extremely security-conscious, and willing to cede many of their natural freedoms to someone who can offer protection. According to Hobbes (in parts of *Leviathan* we did not read), this makes it rational for the contractors to assent to rule by an absolute sovereign, whose will determines the moral code. Anything less would compromise security.\(^3\)

Why does Hobbes think there are no binding moral rules in a State of Nature? His thinking seems to be something like this. Let \(R\) be any putative moral rule.

1. **RATIONAL EGOISM**: \(R\) is binding iff it’s rational (in your interest) to obey \(R\).

2. In the SoN, it’s never rational to obey \(R\).

3. So, \(R\) is not binding in the SoN.

The argument is valid.

The first thing to note about this argument is that Hobbes doesn’t think there is *nothing* to say about what people ought to do in a State of Nature. Hobbes thinks there are lots of things people ought to do in a State of Nature (“force and fraud are in war the two cardinal virtues”). So there are plenty of binding rules—it’s just a stretch to call rules prescribing force and fraud moral.

The second thing to note is that the agreement itself (the envisioned “contract” where people agree to cede many of their natural rights to the Sovereign in exchange for security) is binding on us only insofar as it’s in our interest to accept it. We can flout moral rules whenever it’s in our interest to do so. (This is a rather unpalatable consequence of Rational Egoism.)

\(^3\)Note that while this seems regressive, the idea that political authorities held their powers “in trust” (Kymlicka, p. 187), that rulers had obligations to safeguard the interests of their subjects, was revolutionary, literally.
1.2. THE STATE OF NATURE

1.2.2 John Locke

Locke (1632–1704) rejects the argument: there are binding moral rules in the state of nature. He does this by disputing Rational Egoism: some rules are binding, irrespective of whether it’s in your interest to obey them.

What’s the source of these naturally binding rules? Locke’s thinking seems to be something like this (p. 250). In a nutshell: *ownership is what confers rights* over things (to appropriate/use/harm/destroy them).

1. You have a right over a thing iff you own it (claim)
2. So: you can have rights over a thing iff it’s possible for you to own it (i)
3. It’s possible for you to own a thing only if its nature is subordinate to yours (claim)
4. No one’s nature is subordinate to anyone else’s (people are equal) (claim)
5. So: no one can own anyone else (iii, iv)
6. So: no one can have rights over anyone else (v, ii)
7. God owns everyone (claim)
8. So: no one has rights over anyone else (vi), and God has rights over everyone (vii, i)

Once again the argument is valid. The upshot: by using/harming/killing another person, you do something you have no right to do (and also violate God’s ownership rights over that person). This in turn gives us the right to subject you to coercion, imprisonment, and punishment.

Note that the argument remains powerful even if we omit the stuff about God—so long as we agree that ownership confers rights over things. We tend to accept that people are equal (iv), and it seems plausible that genuine (i.e., legitimate) ownership is impossible among equals. So long as we accept the first premise (ownership confers rights), we’re committed to accepting that no one has rights over anyone else in the State of Nature.

---

4This isn’t to say that Locke is not a contractarian—he is (see especially his *Second Treatise on Government*). The social contract *adds* rights, duties, and rules to the ones that are binding naturally. Contracted laws must respect the natural law—natural law is prior to contracted law.
Note: if Locke supplies a CRITERION OF OWNERSHIP (a way of legitimately appropriating things/acquiring property from our “common endowment”), he can get us even more: natural property rights, which others must respect.

Discussion: what criterion of ownership does Locke supply? Does it seem plausible?

But do rights require ownership? Doubtful. There are many rights that don’t have an obvious origin in ownership. Parents don’t own their children, but they do have rights over them. I have a right to expect my friends to keep their promises, but it isn’t clear how this is an ownership right (and, critically, people who violate it don’t seem to be trying to assert ownership over me).

1.2.3 Jean-Jacques Rousseau

Rousseau (1712–78) disputes premise (ii) of Hobbes’ argument, by appeal to the natural emotion of PITY. Pity is, he says, felt even in the SoN (“a virtue all the more universal and useful to man because it precedes any kind of reflection in him”), and serves as a natural check on people’s malice and tendencies toward violence.

Discussion:

How does pity supply a case against (ii)? (Hint: recall that Hobbes’ argument for the second premise rests on PSYCHOLOGICAL EGOSM. Note that: pity and sympathy might create interests beyond the narrow conception of self-interest Hobbes is using.)

What do you think Rousseau would say about (i)? What do you think he’d say about the conclusion?

This is not to say that Rousseau thinks the SoN is fine, and that no Social Contract is required to justify morality and the state. Quite the opposite—Rousseau’s The Social Contract is the crowning work of the Enlightenment-era Social Contract theorists. Some of its key ideas:

No binding moral rules in SoN. “The passage from the state of nature to the civil state produces a very remarkable change in man, by substituting justice for instinct in his conduct” (Book VIII). Conduct is governed by “physical impulses,” “impulse of appetite.”

There is NATURAL LIBERTY (freedom to govern your action, limited only by
your ability). According to Rousseau, Natural Liberty is a kind of slavery, because it involves rule by the impulse of appetite (Book VIII).

**Social Contract as “alienation” of natural liberty.** Natural liberty is surrendered, in exchange for Civil Liberty (freedom to govern your action, limited by nothing except the General Will (GW), the will of the Body Politic, the collective body that the Social Contract establishes).

Think of the Body Politic not as an actual institution (like a government). For Rousseau, the Body Politic is something different—something like a new identity that replaces the identity that we have in the SoN (as creatures governed by our appetites). The GW is a body of rules that is an expression of that new identity (just as our will in the SoN is the expression of our natural identity).

We retain our appetites and particular desires in the civil state—the expression of which we might call your Particular Will (PW)—but the GW is superior to the PW. Following the PW means being ruled by our appetites and desires. Following the GW means being ruled by Reason.

The conflict of the PW with the GW is Rousseau’s way of dramatizing the conflict between different kinds of identities we can use to govern our conduct (what we might call our animal and civic identities).

**Discussion:**

When someone chooses to follow their PW (and it conflicts with the GW), Rousseau claims it is legitimate to force them to follow the GW. In fact, he claims this “means nothing less than that he will be forced to be free” (Book VII). Why does Rousseau think this? (Hint: the answer is connected to Rousseau’s claim that natural liberty is a kind of slavery.)

**SoN as a normative, rather than historical tool.** Let’s compare / contrast the role that the SoN plays for Rousseau with the role it plays for Hobbes. As with Hobbes, it does seem to be furnishing a normative contrast and a description of the contractors.

**Discussion:**

---

5Rousseau actually calls it a *persona ficta*—a pseudo-person, a legal-political construct, like the modern corporation. The Social Contract is a very confusing book—I think a lot of the confusion can be alleviated by reading it as a metaphor.
How does Rousseau’s description of the contractors differ from Hobbes’? (Hint: what are Rousseau’s contractors concerned to protect that Hobbes’ are not?) How might this manifest in the sort of society Rousseau’s contractors would choose to live in?

It is less clear that it plays a historical role, as it did with Hobbes—that whether the SoN actually happened affects the plausibility of his theory. With Hobbes it clearly did—in view of his Rational Egoism, whether rules are binding depends on how things would go for us if they weren’t in place. If Hobbes is wrong about how the SoN actually was (say, the SoN is pleasant), his account falters. Not so with Rousseau.

This is one reason that modern versions of contractarianism (Rawls, Korsgaard) are best understood on this template.

1.3 Modern Contractarianism

Original versions of contractarianism claimed that moral rules originated from a promise made by a group of contractors (so that our duty to follow the rules follows from our duty to keep our promises). This faces several problems (see Kymlicka’s article):

- No one alive now was party to such a contract. Generally we’re not bound by promises that other people have made.
- Even if we could be bound by the promises of others, there was never an occasion on which people got together, drew up a contract, and agreed on it.

Discussion: what about a hypothetical agreement? Why does Kymlicka find fault with this?

- This leaves an important part of morality (our duty to keep promises) unexplained and unjustified. Other things being equal, we would like like an account which justifies as much of morality as possible. (Note: this applies to Locke’s Natural Rights version of contractarianism too.)

So: a plausible form of contractarianism must appeal to a hypothetical agreement, but cannot try to explain our duty to comply with that agreement as a function of our general duty to keep agreements.

---

6Kymlicka includes Hobbes, Locke, and Rousseau in this group, but we’ve avoided reading their views in this light.
The bindingness of the agreement must have another explanation. As we suggested (§3.1), it is something like this: *a moral rule R is justified because a group of people with qualities I judge important (in deliberating about the rules for human conduct) would agree to R.*

Kymlicka notes two impt. properties of contractarianism, so understood (p. 186).

- **CONVENTIONALISM**: a claim about the validity of a moral rule amounts to a claim about the conventions that a group of people (the contractors) would adopt for themselves
- **INDIVIDUAL CONCERNS**: the description of the contractors must make their concerns salient, in a way that suggests what sort of conventions they would adopt.

The differences between the various versions of contractarianism (Hobbesian, Rousseauian) can, for the most part, be characterized as differences in their descriptions of the situation and concerns of the contractors.

### 1.3.1 John Rawls

Rawls (1921–2002) has an extraordinarily influential (and brilliant) view of the contractors. The situation in which the contract is negotiated is referred to as the **ORIGINAL POSITION**. Contractors in the Original Position:

- Deliberate from behind a **VEIL OF IGNORANCE**: they are made unaware of their social standing, gender, natural abilities, intelligence, and so on (see 1994, p. 363).
- Are self-interested and **RISK-averse**—they are trying to do the best they can for themselves, given the limited information available to them, but they will only agree to a contract which they *know* will leave them better off than any other contract.

What is the point of the Veil of Ignorance? In a word, to make the bargaining situation **FAIR** (Rawls is fond of referring to his view as “Justice as Fairness”). Rawls is supposing that people possess an inherent dignity that *entitles them to “equal consideration” of their interests*, regardless of their social standing, gender, natural abilities, intelligence, and so on.
By placing the contractors under the Veil of Ignorance, Rawls forces them to take seriously the possibility that they could occupy any position in the social arrangement that results from the contract they choose (since they have no idea where they’ll end up).

By making them risk-averse, he forces them to choose a contract in which the least well-off are at least as well-off as they are under any other contract (the **Difference Principle**; p. 365).

- **Discussion:**
  - Is being risk-averse in this way rational or irrational? Are you risk-averse in this way?
  - What conceptions of social justice does the Difference Principle rule out? (Hint: think of a society in which capital is distributed in such a way as to maximize GDP.) Are inequalities of resources ever justifiable on this view?

It’s interesting to think about the similarities between Justice as Fairness and Rousseau’s conception of Justice as the General Will. Rawls, like Rousseau, views the rules that would be agreed to in the Original Position as “self-imposed” (p. 364); the rules binding on people in a society organized along Rawlsian lines are, in a real sense, *laws that they give to themselves*.

### 1.3.2 David Gauthier

Gauthier (1932–) tries to defend the orthodox Hobbesian view using modern tools (in particular, tools from game theory, like the Prisoner’s Dilemma). Like Hobbes, he presupposes Rational Egoism—moral rules are justifiable (binding) iff it is in your interest to obey them.

The idea, as with Hobbes, is that moral rules are constraints on “maximizing choice” (behavior that maximizes your expected utility) that are justified because they lead to a **Pareto Optimal** state: “no one can do better without someone doing worse” (p. 367). In the Prisoner’s Dilemma, the state where both cooperate is Pareto Optimal, and the Equilibrium state is Pareto Suboptimal.

A Pareto Optimal state is not necessarily optimal for you: you could do better (by defecting while your fellow player cooperates). Given that this outcome isn’t really available to you, however, you both do best by agreeing to constrain your behavior, but only conditional on the other player doing this as well, so that you both cooperate. “If one benefits more from a constraint on others than one
loses by being constrained oneself, one may have reason to accept a practice requiring everyone, including oneself, to exhibit such a constraint” (p. 367). The correct moral code is, Gauthier suggests, just a collection of such practices.

- **Discussion:**
  
  Why might the existence of such a practice to make defecting even more attractive? (Hint: does the probability of your fellow player cooperating increase or decrease with such a practice?)

  Given this, how can it nevertheless be rational for you to abide by the constraint? What do you lose by having a reputation as untrustworthy, or as a “straightforward maximizer”? What do you gain by having a reputation as a “constrained maximizer”? (See Gauthier’s Jones/Smith case, p. 369.\(^7\))

  What sorts of social conditions would you expect to enhance cooperation among individuals?\(^8\) (Hint: what sorts of conditions would you expect to encourage people to TRUST each other?)

So far, so Hobbesian (although the game theoretic elements of Gauthier’s account are a definite improvement over Hobbes’ sparse Rational Egoism).

Gauthier’s account is, so far, recognizably a contract account: a moral rule \( R \) is justifiable to \( A \) iff \( A \) would hypothetically agree to a society governed by \( R \) in an appropriate “pre-moral” situation (namely: the situation where everyone is trying to negotiate a set of rules that makes things go as well as they can, overall, for her).

Gauthier takes notice of a problem for Hobbes’ account (which we’ll discuss more later on): the rules you’ll agree to (rights you negotiate for yourself) in such a situation “depend in large part on [your] negotiating position vis-à-vis [your] fellows” (p. 372). In particular, the power relations that are in place in your society will affect your negotiating position, and, so, the rules you’ll agree to. The rules negotiated by people in weak bargaining positions will not be as beneficial for them as the rules negotiated by people in strong bargaining positions. But that seems wrong.

- Gauthier tries to get around this by insisting that there is something “arbitrary” about allowing existing power relations to determine negotiating

---

\(^7\)The Jones/Smith case is an example of a more general problem: if you know the number of future games you’ll be a player in (in this case, two), cooperating is irrational.

\(^8\)Robert D. Putnam’s Making Democracy Work is a great exploration of this question.
CHAPTER 1. CONTRACTARIAN ETHICS

positions. Instead, the appropriate pre-moral situation is one in which existing power relations are erased, and people negotiate from positions of equal standing (pp. 372-3). (Gauthier isn’t explicit about this, but such a situation might look a lot like Rawls’ Original Position.)

• Discussion:

Is there any way for Gauthier to justify appealing to a “fair” (non-arbitrary) negotiating situation, or is his so-called Hobbesian contractarianism just a version of Rawlsian contractarianism?

Maybe. Consider Gauthier’s (true) claim that existing power relations reflect a society’s existing moral code (p. 372). So in allowing existing power relations determine negotiating position, we haven’t really reached a pre-moral situation—we’re really allowing the existing moral code to determine aspects of the justified moral code.

1.4 Critiquing Contractarianism

Kymlicka’s (1993) and Midgley’s (1994) articles raise two worries for the contractarian answer to the Task of Justification.

Worry 1. It seems like a description of a typical contractor is doing all the theoretical work by itself. Contractors are just idealized versions of ourselves (Kymlicka, pp. 193-4). Why is this a problem?

• The Task of Justification is shifted, not resolved. Instead of the question why be moral?, our question is why care about the preferences of the contractors? If we can’t answer the second question, we haven’t answered the first.

Is this asking too much from a moral theory? Maybe—it seems we can make progress on the Task of Justification, even without completely answering it. We could, for instance, bring someone to see how her commitment to the concerns that the contractors have generates a commitment to other concerns. Maybe this is good enough—it’s enough for a moral theory to justify itself to people who share certain ground-level commitments with us.

• Instead of looking to the result of an agreement between the contractors, we can often enough just look directly at their preferences (the content of the preferences and of the agreement will coincide, if the contractors are idealized enough). The contract part of the theory doesn’t seem to be essential.
1.4. CRITIQUING CONTRACTARIANISM

- **Discussion:** how might Hobbesian contractarianism claim to avoid these worries?

**Worry 2.** According to contractarianism, claims Midgley, “[A]ll significant moral relations between individuals are the symmetrical ones expressed by contract” (1994, p. 376). But then how are rights secured for those who cannot be party to the social contract (because physical or mental limitations prevent them from being party to any contract (infants, the mentally incapacitated, non-human animals, etc.) or for entities that cannot even sensibly enter into a contract (families, nations, etc.)? (See Midgley (1994, pp. 381-2) for an impressive list of things the contractarian has difficulty generating duties toward.)

- Hobbes has a very hard time with this: the rights an individual has are precisely those she’d negotiate for herself in a SoN. As Kymlicka says, “Rights depend on one’s bargaining power” (1994, p. 189). More precisely: the duties A recognizes herself as having to B depend on B’s bargaining position with respect to A. So it is hard to see how Hobbes can even get *symmetrical* moral relations out of a social contract, since power relations (hence bargaining positions) tend to be asymmetric.

- As does Rousseau, for whom the moral community is the Body Politic: Civil Liberty is acquired in exchange for cession of one’s Natural Liberty, which establishes membership in the Body Politic.

- **Discussion:** is this a problem for Rawls’ contractarianism? Gauthier’s?

We might think that Rawlsian contractarianism has no problem with Worry 2: we might just require that the contractors care about the interests of infants, the mentally handicapped, and non-human animals. Rawls has, after all, already modified the description of the contractors so that they embody some of our most fundamental moral concerns (impartiality and fairness). If this is okay, why not gild the lily?

- In support of this, note that contractarians face the following dilemma: either we restrict the description of the contractors, so that they embody some of our fundamental moral concerns, or accept, for instance, that existing power relations determine, in part, what’s right and wrong.

  The latter way is clearly unacceptable, so we should really opt for the former. But once we’ve done that, there’s no barrier, in principle, to building some more things in.
• Midgley’s response (echoed by Kymlicka): “As far as importance goes, it is certainly possible to argue that some of these sorts of beings should concern us more and others less: we need a priority system. But, to build it, moral arguments are required” (p. 382).

In other words, giving the contractors a concern for X presupposes that concern for X is morally important (or that we have a duty to promote the wellbeing of X). There is a host of meaty moral questions (like whether X is morally important) that fall outside the scope of a contractarian justification.

How devastating is this point? It depends on what we take the goal of the contractarian theory to be. Contract ethics could be understood in each of the following guises:

• An attempt to identify what is fundamental to our moral practice, to answer the question: what set of ground-level concerns generate/ground the whole of everyday morality?

Midgley and Kymlicka appear to understand contractarianism in this way. They may be right that contractarianism cannot work under this guise—some of morality cannot be grounded on the agreement of rational individuals.

• An attempt to identify a procedure whose outcome is a morality that is binding on us, outside of which (“pre-procedure”) there can be no binding morality—a procedure, in other words, that generates the correct moral code.9

Discussion: do Midgley and Kymlicka’s criticisms apply to this kind of contractarianism?

• An attempt to resolve the Task of Justification—to explain how various important duties to each other can be grounded on more basic, shared moral commitments.

This avoids the criticisms. It is hardly a strike against this kind of contractarianism that only rational persons can be contractors, because rational persons are the ones to whom we are trying to justify morality.

---

9As we’ll see next class, something like this is actually Rawls’ view, although it is not required by his idea of the Original Position.
Chapter 2

Consequentialist Ethics

2.1 Consequentialism, In General

As Pettit notes (1993, pp. 230-1), consequentialist theories are distinguished by these traits.

A substantive conception of intrinsic goods. Every consequentialist theory endorses a proposal along the following lines: $X$ is intrinsically good iff $X$ has property $P$. (Things are extrinsically good insofar as they promote the existence of the intrinsic goods.)

- What differentiates consequentialist theories from one another is different substantive conceptions of the intrinsic good.
- Utilitarians will identify intrinsic goodness with happiness (impersonally considered).
- States of affairs (descriptions of the world at a point in time) are good insofar as they contain instances of the good. The goodness of a state of affairs is a function of whether or not there are good things in it. States of affairs can be ranked (ordered) according how good they are.

Teleology. An act’s rightness or wrongness is a function of the goodness (and badness) of its consequences. This is teleological because it involves identifying an end (the good) that’s to be promoted. Deciding whether or not to do the act should only be a matter of how that act promotes the good. Consequentialist theories are essentially forward-looking. What they endorse is bringing about the best states of affairs.
“Consequentialism is the view that whatever values an individual or institutional agent adopts, the proper response to those values is to promote them” (1993, p. 231).

• Teleological views are contrasted with ATELEOLOGICAL views. According to ateleological views, what’s important to evaluating rightness and wrongness is the nature of the act itself (not its consequences) or the motives of the act’s agent.

“Opponents of consequentialism see the relation between values and agents as a non-instrumental one: agents are required or at least allowed to let their actions exemplify a designated value [i.e., side-constraint], even if this makes for a lesser realization of the value overall” (ibid.).

Note how contrary this is to consequentialism, which says that how an act is performed is irrelevant; what matters about actions are features extrinsic to the act, i.e., its consequences.

– Moderate deontology (W.D. Ross). Consequences should sometimes make a difference in determining what your ethical duties are.

– Absolute deontology (Immanuel Kant). Consequences never make a difference in determining what your ethical duties are. So, if it’s your duty to do X, then that duty holds, no matter if doing X brings about the destruction of the world.

• ⇒ Formal analysis of rightness. An act X is morally right for an agent A iff there is no alternative Y to X such that A could have done Y and Y brings about a better state of affairs (outcome) than X. In brief, an act is right iff it brings about a better outcome than any other action the agent could have done.¹

Absence of side-constraints. A side-constraint view holds that value can be promoted only on the assumption that a certain constraint is fulfilled. Any promotion of value which involves violation of the constraint is ipso facto impermissible. (Pettit refers to taking this response toward the value required by the side-constraint as “HONORING” that value, which is distinguished from a response of “SEEKING TO PROMOTE” that value.) Obviously consequentialists cannot endorse any sort of side-constraint.

¹We can, and Pettit does (see 1993, p. 233), reformulate this analysis to encourage maximization of expected good, so that if X offers a certainty of saving 40 lives and Y offers a 50% chance of saving 100, the theory requires an agent to do Y.
A consequence of consequentialism is that individuals lack inviolable rights. Individuals aren’t ends in themselves; according to the theory, individuals are means to the end of value-promotion. If individuals have worth, it’s derivative worth – individuals are valuable only insofar as their existence promotes the good.

Trolley case.

Consequentialism as the simpler theory. Nonconsequentialists, according to Pettit (1993, pp. 238-9), sanction two kinds of response toward values: both honoring and promotion. Take, for instance, rights. Certain nonconsequentialists think rights supply side-constraints (they should be honored). But rights are also valuable (they should be promoted, within the boundaries supplied by side-constraints). Consequentialists sanction only one kind of response toward values: the response of promoting them.

Pettit is worried that honoring a value is not a rational response to it. We normally understand an agent to be rational insofar as she selects action that maximize expected value; why would a moral theory be different? (Nozick, we will see, gives an argument that it is rational for a moral theory to honor side-constraints, because it is rational for agents to maximize expected utility while respecting side-constraints.)

2.2 Consequentialism, In Particular: Utilitarianism

According to Mill, utilitarianism is defined as:

“The creed which accepts as the foundation of morals “Utility” or the “Greatest-Happiness Principle,” holds that actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness. By happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure.”

In other words, an act $X$ is right iff $X$ maximizes the total amount of pleasure in the world and minimizes the total amount of pain in the world.

Mill’s utilitarianism is a version of consequentialism. It’s hedonistic (the only intrinsically good things are pleasure and pain) and universalistic (the only intrinsically good things are pleasure and pain, considered impartially; that’s to
CHAPTER 2. CONSEQUENTIALIST ETHICS

say, it doesn’t matter whose pleasure and pain; everyone’s counts the same, including that of animals, though with some qualifications). Mill’s hedonism and universalism together give us his full conception of what is intrinsically good: the good is identified with the pleasure and the absence of pain, of no one at all in particular.

According to the greatest happiness principle... the ultimate end, with reference to and for the sake of which all other things are desirable—whether we are considering our own good or that of other people—is an existence exempt as far as possible from pain, and as rich as possible in enjoyments, both in point of quantity and quality.

2.3 An Objection to Utilitarianism

A “philosophy of swine.” Should the pleasures of a pig and the intellectual pleasure of reading philosophy should be accorded equal weight by an ethical theory, if they bring about similar quantities of pleasure in the individuals who experience them? Are the pleasures of the sadistic torturer an intrinsic good, on a par with a pleasure of similar magnitude derived from reading philosophy?

Mill’s response. Indeed, that’s implausible. But he doesn’t say it! A pleasure’s value isn’t just a matter of how good it feels. That’s to say, pleasures aren’t ordered just by how good they feel.

• Experienced judges. When trying to rank two pleasures $X$ and $Y$, one asks people who have experienced $X$ and $Y$ to rank them. If there’s a general preference for $X$ over $Y$, then $X$ should be ranked ahead of $Y$.

• According to Mill, experienced judges will agree that intellectual pleasures are superior to pleasures of swine.

• Lexical priority. If experienced judges agree that $X > Y$, then $X$ always takes precedence to $Y$. No quantity of $Y$ should ever be taken in exchange for any quantity (no matter how small) of $X$.

• Discussion: does Mill endorse lexical priority? Or is the preference given to intellectual pleasures defeasible (oughtweigh-able)?
2.4. AN OBJECTION TO CONSEQUENTIALISM

**Worry.** Is it fair to privilege the opinions of experienced judges in this way? In effect, doesn’t this amount to privileging the aesthetic preferences of the aristocratic (who uniformly prefer the sorts of pleasures only they have had the benefit of enjoying) over the privileges of the masses? (Can you think of some examples?)

- We could respond by ditching the experienced judges test, in favor of liberal utilitarianism. We simply rank desires according to their strength (how much the individual wants the thing in question). But this leads back to the problem that Mill was trying to deal with (and also commits us to giving consideration—even preference—to the acute desires of sadists).

2.4 An Objection to Consequentialism

**Counterintuitive ethical commitments.** For example.

- Consequentialism sanctions (indeed, requires) “horrendous deeds, so long as they promise the best consequences” (1993, p. 234).

- Consequentialism demands that our moral deliberation just involve weighting the value of one course of action against the value of another, an “incessant practical casuistry.” This demands an objectionable revision of the way we actually deliberate about moral questions (see Pettit’s discussion, 1993, pp. 234-5).

  - Consequentialism is forward looking. So obligations incurred in the past (e.g., to keep your promises) are always renegotiable, if it turns out that violating them will bring about a better state of affairs. But ordinarily we don’t think that morality allows promises to be so easily ignored.

  - Consequentialism has problem making room for inalienable rights. But ordinarily we think that everyone has a right against torture, and we rely on such intuitions in our ordinary deliberation about what to do. How can consequentialism make room for these sorts of rights? How can it make room for the way that people privilege these rights in ordinary deliberation?

**Some consequentialist responses.** (This draws on Pettit, 1993, pp. 235-7.)

- *Horrendous deeds only in horrendous circumstances.* As Pettit notes, torture is awful, but so is allowing a bomb to detonate because you are unwilling to torture.
• *Restrictive consequentialism*. Consequentialism does not have to be a theory of how we ought to deliberate; it can just be a theory of the right and the good. According to restrictive consequentialism, for example, there are justified rules of action (rules of thumb, e.g., always keep your promises), justified motives (for things besides the good) that we ought to structure our deliberation in accordance with. The existence of the rules of thumb and of motives for things besides the good is justified because deliberating in accordance with these rules and motives ultimately leads to the best consequences. In short: people making things besides the good their motive is justified because it ultimately promotes the good.
Chapter 3
Deontological Ethics

3.1 Overview

Consequentialism. An act $X$ by $A$ is right iff $X$ brings about an outcome that’s at least as good as any other outcome that $A$ could have brought about.

- Consequentialists care about the extrinsic features (consequences) of action.

Deontology. An act $X$ by $A$ is right iff the $X$ has intrinsic feature $F$ or if $X$ is caused by a motive of $A$ with intrinsic feature $G$.

- You get different deontologies by filling in different $F$’s and $G$’s.

3.2 Kantian Ethics

Immanuel Kant (1724–1804) in Brief. An act $X$ done by $A$ is right iff $A$’s maxim of action passes the Categorical Imperative Test.

Note: Kant evaluates actions by looking at their maxims—a general law of actions that the agent was utilizing in acting—and nothing else.

Kant endorses a version of Absolute Deontology. Consequences never make a difference in determining what your ethical duties are. So, if it’s your duty to do $X$, then that duty holds, no matter if doing $X$ brings about the destruction of the world.
3.2.1 Reason, Freedom, and The Good Will

Reason and Freedom. Kant views the Moral Law as being given by your own reason, and that any action that violates the Moral Law is an action against your own reason. Denying the Moral Law is rationally tantamount to accepting a contradiction. The Moral Law is organic to a rational will, in a special sense.

“The essential point in all determination of the will through the moral law is this: as a free will, and thus not only without co-operating with sensuous impulses but even rejecting all of them and checking all inclinations so far as they could be antagonistic to the law, it is determined merely by the law ... Freedom...consists, however, only in the fact that it limits all inclinations, including self-esteem, to the condition of obedience to its pure law.” (1994, pp. 130-1).

Good Will. The only thing that is unconditionally good (i.e., intrinsically valuable, without qualification, for its own sake) is a good will (1994, p. 123). Other aspects of the agent may be good, but only conditionally or contingently so—in short, certain conditions can make them bad.

A GOOD WILL is a will that obeys the Moral Law, simply because it is the Moral Law, not for any other reason. It is not sufficient to obey the Moral Law, if the obedience is done for the wrong reason. Kant’s example of the selfish, but honest, shopkeeper (see 1994, p. 126 and 1993, p. 177).

• In order to have MORAL WORTH, an action must obey the moral law and originate from a Good Will. (In other words, the motives of the agent, rather than the nature of the action, is what Kant think the moral theorist ought to be interested in.)

“If the determination of the will occurs in accordance with the moral law but only by means of a feeling of any kind whatsoever, which must be presupposed in order that the law may become a determining ground of the will, and if the action thus occurs not for the sake of the law, it has legality but not morality” (1994, p. 129).

 “[The moral value of an action] does not depend on the reality of the object of the action [i.e., whether it realizes the goal it was undertaken to achieve] but merely on the principle of volition [maxim, motive] by which the action is done without any regard to the objects of the faculty of desire [desired outcomes, the goals of the agent]” (1994, p. 128).
3.2. KANTIAN ETHICS

Later: “Thus the moral worth of an action does not lie in the effect which is expected from it or in any principle of action which has to borrow its motive from this expected effect. For all these effects ... could be brought about through other causes and would not require the will of a rational being” (1994, p. 129).

So obviously Kant is not a consequentialist.

- As we’ll see below, Kant says that an action is right (obeys the Moral Law) iff it passes the Categorical Imperative Test. So: a will is not good, unless the maxim that it acts on passes the Categorical Imperative Test, and the agent does the action the maxim recommends because it passes the Categorical Imperative Test.

3.2.2 ‘The Formula of Universal Law’

Maxims. Maxims are general laws of action, having the following form: In circumstances C, do action X in order to obtain an outcome O.

Example: If I need money (C), then I will obtain a loan and make a lying promise to repay the loan (X) in order to obtain money (O).

Categorical Imperatives, in general. These are commands to do something, no matter what you want. (To be contrasted with a hypothetical imperative, which is a conditional or contingent command—do X, if you want Y and believe that X-ing will get you Y.)

- Kant thinks that the Moral Law is given in the form of categorical imperatives. So, your moral duties are never a function of what you want; the moral law is categorical, which is to say that it doesn’t care what you want!

‘The’ Categorical Imperative. ‘The’ Categorical Imperative = The Moral Law.

“So act that your principle of action [maxim] might safely be made a law for the whole world.”

In other words, act only on those maxims that you can consistently will to be universal laws of nature. Roughly, acting on a maxim M is permissible only if you can consistently will that everyone should make M a rule of their action. Permissible maxims can be universalized.
CHAPTER 3. DEONTOLOGICAL ETHICS

Testing for universalizability. To test whether or not a maxim \( M \) (in \( C \), do \( X \) to get \( O \)) is allowed by CI, imagine a world in which \( M \) is a law of action for everyone (so, whenever anyone’s in \( C \), they do \( X \) to get \( O \)). If it’s still possible to follow \( M \) in that world, then it’s permissible to make \( M \) a law of your action in this world.\(^1\)

- Example: Let your maxim \( M \) be: “if I need money, then I will obtain a loan and make a lying promise to repay the loan in order to obtain money.” Can \( M \) be universalized?

No. Imagine a world in which everyone acts on \( M \). In such a world, the practice of lending breaks down, so it’s not possible to follow \( M \) in such a world.

In general, *promises should always be kept*, the maxim you use in breaking a promise can’t be universalized. If you tried to universalize the maxim, the institution of promise-making would cease to exist, and so it would be impossible to follow the maxim in a world in which the maxim were a law of nature.

This gives the sense in which Kant is an absolute deontologist. Even if you could do a great good by telling a lie or by breaking a promise, it’s morally impermissible for you to do so. The Moral Law is rigid; the consequences just don’t matter.

Kantian Procedural Justice. Why might you view the CI Test as a version of pure procedural justice? (Hint: see O’Neill, 1993, p. 176, in particular her remark that “Kant’s central move is to construct the principles of ethics according to rational procedures.”) Why does Kant (and, for that matter, Rawls) deny that “the principles of good willing can be fixed by reference to an objective good or *telos* at which they aim.”

3.2.3 ‘The Formula of the End in Itself’

Interestingly, Kant offers a different formulation of the Moral Law that he claims is equivalent to the Categorical Imperative.

Kant thinks that the only intrinsically valuable thing is a good will. From this, Kant thinks it follows that human beings (insofar as they approximate a good will, even if only a little bit) are intrinsically valuable. From which it follows that humans have intrinsic worth that cannot be infringed for any reason.

\(^1\)My understanding of O’Neill’s formulation of the CI Test. See (1993, p. 177).
3.2. KANTIAN ETHICS

- Discussion: is this a valid argument?

According to Kant, this means that it’s never permissible to use another person solely as a means to your own ends (i.e., solely as something having only instrumental worth, vis-à-vis your ends or goals). This leads him to endorse a second formulation of the Categorical Imperative, as follows.

“So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means.”

According to Kant, it is permissible to use other people as means to your ends, but only if, in so doing, you simultaneously respect their humanity / their rational autonomy / their ability to govern their actions as free creatures / their ability to consent.

Kant thinks that this formulation is equivalent to (i.e., yields the same results as) his earlier formulation of the Moral Law.

- Discussion: is this actually the case? (Perhaps: Kant probably thinks that being truly free—a lawgiver to yourself—implies valuing rational autonomy per se. Kant’s thinking is probably something like: you value rational autonomy per se iff you follow the Formula of the End in Itself iff you follow the Formula of Universal Law.)

3.2.4 Empty Formalism?

According to Hegel (1770–1831), Kant reduces ethics to “empty formalism.” (For a short, but smart, discussion, see O’Neill, 1993, pp. 181-2.)

“If the definition of duty is taken to be the absence of contradiction, formal correspondence with itself—which is nothing but abstract indeterminacy stabilized—then no transition is possible to the specification of particular duties nor, if some such particular content for acting comes under consideration, is there any criterion in that principle for deciding whether it is or is not a duty. On the contrary, by this means, any wrong or immoral line of conduct may be justified” (1994, p. 132).

The Categorical Imperative Test, according to Hegel, gives no more substantive ethical advice than: do not contradict yourself:
“But if [and only if!] it is already established on other grounds and pre-supposed that property and human life are to exist and be respected, then indeed it is a contradiction to commit theft or murder; a contradiction must be a contradiction of something, i.e. of some content presupposed from the start as a fixed principle” (ibid.)

Discussion: what sort of conditions does Hegel think need to be in place to have a substantive notion of right and wrong? For genuine freedom of the will? (Hint: see pp. 133-4).

3.3 Agent-Relativity and Agent-Neutrality of Value

“Tell me honestly, I challenge you—answer me: imagine that you are charged with building the edifice of human destiny, the ultimate aim of which is to bring people happiness, to give them peace and contentment at last, but that in order to achieve this it is essential and unavoidable to torture just one little speck of creation, that same little child beating her chest with her little fists, and imagine that this edifice, has to be erected on her unexpiated tears. Would you agree to be the architect under those conditions? Tell me honestly!”

Last time, we considered two objections to consequentialism.

- Consequentialism sanctions (indeed, requires) “horrendous deeds, so long as they promise the best consequences” (1993, p. 234).

- Consequentialism demands that our moral deliberation just involve weighting the value of one course of action against the value of another, an “incessant practical casuistry.” This demands an objectionable revision of the way we actually deliberate about moral questions (Pettit’s discussion, 1993, pp. 234-5).

Today, we’ll consider two new, but related, objections.

3.3.1 John Rawls

Rawls’ argument against consequentialism asks us to consider different ways of distributing intrinsic value among a group of people: $D_1, ..., D_n$. Let $U(D_i)$ be the total value in $D_i$.

1. If consequentialism is true, then $D_i ≻ D_j$ iff $U(D_i) > U(D_j)$.

2. Whether $U(D_i) > U(D_j)$ depends only on the total value in $D_i$ and $D_j$, not at all on how value is distributed among separate individuals.
3. But whether $D_i \succ D_j$ does depend on how value is distributed among individuals. (According to Rawls, consequentialism “does not take seriously the distinction between persons.” It treats the job of ethics to maximize the happiness of a composite “person” that just aggregates the wants of everyone of society. In aggregating people’s wants in this way, we lose the ability to make distinctions that depend on distinguishing one person from another person—whether an individual deserves a certain allocation of value, whether it is fair to that individual, etc. And that is wrong.)

4. So, consequentialism is false.

A wonderful, and deservedly famous, quote:

“[According to consequentialism], [S]eparate individuals are thought of as so many different lines along which rights and duties are to be assigned and scarce means of satisfaction allocated...so as to give the greatest fulfillment of wants” (1994, p. 338).

### 3.3.2 Bernard Williams

Williams describes two cases, similar in all the relevant respects (we’ll focus on “Jim and the Indians”). The relevant features of the case (rows are actions Jim can take, columns describe the outcome of those actions):

<table>
<thead>
<tr>
<th></th>
<th>Dead</th>
<th>Alive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shoot one</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>Do nothing</td>
<td>20</td>
<td>0</td>
</tr>
</tbody>
</table>

Take any formulation of consequentialism which places intrinsic value on human life / happiness / whatever. That form of consequentialism will claim that:

- Jim ought to shoot one of the Indians.
- The fact that shooting one will save nineteen is a reason for Jim to shoot one.
- The fact that it is Jim who will be the one shooting the one Indian is NOT a reason for Jim not to shoot one.

Williams graciously grants the first two of these things to the consequentialist. What he objects to is the third: in general, the fact that I will be killing someone is often a moral reason for me not to do it. This is a familiar objection to consequentialism (which we’ll encounter again in Ross): consequentialism doesn’t seem to allow certain facts to count as moral reasons which, it seems,
CHAPTER 3. DEONTOLOGICAL ETHICS

ought to count as moral reasons (e.g., the fact that doing some action will compromise *my* integrity, as Williams puts it on p. 340). Consequentialism requires that we assign less weight “to our own avoidance of wrongdoing than we do to the avoidance of wrongdoing tout court, or the prevention of wrongdoing by others” (Davis, 1993, p. 207).

**Agent-Relativity and Agent-Neutrality.** Williams, in this piece, is making an extended plea for the recognition of a class of Agent-Relative, rather than Agent-Neutral, moral reasons. To illustrate, consider the difference between the following competing justifications for some agent $B$ keeping a promise to her friend $A$.

- **Reason 1:** keeping the promise will make someone happy
- **Reason 2:** keeping the promise will make $B$’s friend happy
- **Reason 3:** keeping the promise means that $B$ respects her relationship to her friend

Reason 1 is agent-neutral: whether it is a reason for some agent does not depend on *who the agent is* or *what moral relationships the agent partakes in*. Reasons 2 and 3 are agent-relative (assuming that their status as moral reasons is *sui generis*—doesn’t depend on their promoting or being conducive to the Good). Whether they apply depends on whether the agent has a friend whom keeping the promise will make happy, or whether the agent has a friendship that she can respect by so doing.

Specifically, in this piece, Williams is focusing on the idea that *an agent’s own projects* cannot supply any sort of agent-relative reason in favor of fulfilling them, if consequentialism is correct.

> “[The moral agent] is the agent of the satisfaction system who happens to be at a particular point at a particular time... His own decisions as a utilitarian agent are a function of all the satisfactions which he can affect from where he is: and this means that the projects of others, to an indeterminately great extent, determine his decision” (1994, p. 344)

That is to say, according to consequentialism:

- The fact that $X$ fulfills one of $A$’s fundamental or ground-level projects cannot be a sui generis moral reason for $A$ to $X$. 
The fact that $Y$ conflicts with one of $A$’s fundamental or ground-level projects cannot be a sui generis moral reason for $A$ not to $X$.

That’s because consequentialism, in general, has an exceptionally hard time making room for agent-relative reasons. That is problematic: it alienates us from our most basic commitments, projects, and concerns.

“It is absurd to demand of such a man, when the sums come in from the utility network which the projects of others have in part determined, that he should just step aside from his own project and decision... It is to alienate him in a real sense from his actions and the source of his action in his own convictions. It is to make him into a channel between the input of everyone’s projects...and an output of optimific decision” (1994, p. 345)

Because consequentialism regards moral agents as mere channels between the circumstances and an optimal outcome (and disregards the intrinsic moral significance of their own projects), it is too demanding a moral theory. A moral theory is too demanding if it claims that an agent’s projects—for instance, the fact that doing $X$ would require violating them—cannot supply a sui generis moral reason.

Note: we need to distinguish this objection from its less plausible relative: the objection that consequentialism does not allow agents to care about their own projects, that it requires them to adopt as their “ultimate project” the promotion of the good. This is not necessarily the case: according to consequentialism, it is right for an agent to have her own projects, care deeply about them, even adopt them as ultimate or intrinsic goals, so long as having that sort of psychology is what promotes the most good. This is not the objection that Williams is making.

**Doing vs. Allowing.** Williams’ discussion of integrity and agent-relative reasons is closely related to the deontological distinction between doing and allowing.\(^2\) For the agent-relative fact that doing $X$ would require that $^\ast A $ kill someone to provide a sui generis reason for $A$ not to $X$, we need to (i) make a distinction between what $A$ does intentionally and what $A$ allows or causes to happen (without intending for it to happen)\(^3\) (ii) claim that this distinction can have some moral significance.

\(^2\)See also Davis (1993, pp. 209-10).

\(^3\)Roughly equivalently, we need to recognize actions (or motives), and not just outcomes (consequences), as objects of moral evaluation.
• If we failed to recognize this distinction, then the presence or absence of intentions on the part of A would make no difference to the rightness or wrongness of X. The only features of the circumstances that could matter to A’s decision—that could figure in a description of A’s moral reasons—are features of the action impersonally described (namely, what will happen as a result). The fact that some outcome happens as the result of X is an agent-neutral fact. (We should discuss why this is the case.)

• Agent-relative moral reasons (of the sort Williams has in mind) require distinguishing between what an agent does and what happens as a result of what an agent does.

3.4 Prima Facie Duties

In “The Personal Character of Duty,” W. D. Ross’ general target is any attempt “to base rightness on productivity of some sort of result” (1994, pp. 332-3). Ross denies that “productivity of the maximum good” is what “makes all right actions right” (p. 333). Which is to say: he denies teleology.

Why does he do so? Consequentialism concerns itself with the promotion of some good, and, therefore, with future outcomes. But Ross notes that we often take facts about the past, rather than the future, to be relevant in moral deliberation (see also Dancy, 1993, pp. 220-1). For instance, the fact that I have made a promise is generally what supplies the Moral Reason for thinking that I shouldn’t break my promise. Consequentialism cannot explain why a fact about the past can supply this sort of moral reason.

Why take the fact that I have made a promise to supply a sui generis moral reason? The argument is simple (see p. 334).

(3.1) If the only moral reasons we had were facts about future consequences, then promise-breaking would be permitted even when the future consequences of promise-breaking were just slightly better than those of promise-keeping.

(3.2) Promise-breaking is NOT permitted when the future consequences of promise-breaking are just slightly better than those of promise-keeping.

(3.3) So, we have moral reasons to keep promises that are not facts about future consequences.

What’s the nature of sui generis moral reasons? Ross thinks that they are supplied by what he calls Prima Facie Duties (PFDs): the fact that X violates a PFD supplies a (not necessarily decisive) reason against doing X, while the fact
3.4. PRIMA FACIE DUTIES

that X respects a PFD supplies a (not necessarily decisive) reason for doing X. PFDs are “not strictly speaking duties, but things that tend to be our duty” (p. 334).

Some examples of PFDs (1994, p. 336):

1. Relieve distress / promote happiness when possible (beneficence).
2. Make reparation for past wrongs / render gratitude for past benefits.
3. Distribute benefits according to desert / fairness.
4. Keep your promises / do not be dishonest.

Prima facie duties have the following features (for discussion, see especially Dancy, 1993).

- They’re prima facie! The fact that X violates a PFD supplies a reason against doing X, but it can still be permissible to do X. The reasons supplied by PFDs are inconclusive. (That’s good, because....)
- They can conflict, depending on the circumstances.
- They’re all basic—there is no “coherence,” no unifying moral principle or value that PFDs can be seen to embody or can be derived from. It follows that they’re all on equal footing—no PFD is more basic than any other. That’s to say, there is no ultimate PFD such that all the other PFDs can be derived from it and justified in virtue of their connection to it.
- They somehow determine Ultima Facie (or “all things considered” / sans phrase) duties. Your ultima facie duty in circumstances C is a matter of considering (and somehow weighting) the reasons supplied by the PFDs that give a verdict in C. In general, you try to follow the PFD (or group of PFDs) that is, in C, “more of a duty” than the PFDs that give a conflicting verdict in C.

When I am in a situation, as perhaps I always am, in which more than one of these prima facie duties is incumbent on me, what I have to do is study the situation as fully as I can until I form the considered opinion ... that in the circumstances one of them is more incumbent than any other: then I am bound to think that to do this prima facie duty is my duty sans phrase in the situation.
How much of a duty something is, can, and often does, depend on the circumstances.

- **Value Pluralism.** There are lots of different things we ought to care about. Moreover, the proper response to some is one of seeking to promote, while others ought to be honored or respected.

So Ross, in contrast with Kant, is a **moderate deontologist** (Kant is an absolute deontologist). The consequences can sometimes justify breaking a promise.

- How does this work? Some PFDs tell us to care about consequences, and we can imagine circumstances where the future-oriented duty to relieve distress, for instance, becomes more important than our duty to keep a promise.

**Objections and Replies.**

- **Objection:** This is a very complicated theory.
  
  **Reply:** “it is more important that our theory fit the facts than that it be simple” (1994, p. 334).

- **Objection:** A list of a bunch of normative platitudes isn’t a normative theory at all.
  
  **Reply:** (i) sure it is; the ethic of PFDs provides a distinctive way of looking at ethical questions; (ii) this is how our moral practice actually looks—we have defeasible rules, without any underlying theoretical rationale.

- **Objection:** It is a bit teleological—we are trying to maximize fulfillment of PFDs (or PFDs, as weighted by their importance in a set of circumstances).
  
  **Reply:** Sure, it’s teleological in this attenuated sense (and probably this is true of any deontological theory). But not in a more interesting sense:

  - Our duties can depend on the moral relationships that we have partaken in (being a friend to A, having made a promise to A, etc.) (see 1994, p. 335). Moral reasons can be Agent-relative, for short. Agent-relativity of value is what Ross is getting at when he talks about the “personal character of duty”.

---

4For your consideration: does this fact about an ethic of prima facie duties—the fact that it can be construed as an ethic of value-promotion—make it immune to Pettit’s objection to honoring, as opposed to promoting, values?
3.4. PRIMA FACIE DUTIES

There is no general procedure for weighing PFDs in a circumstance (we do not, for instance, weight them according to how much of the good they’ll result in). The agent simply examines the circumstance and makes a considered moral judgment (which Ross and Dancy would style an application of her intuition) about it. Post hoc, it’s possible to rationalize it in this teleological way—we can say the agent was trying to maximize fulfillment of PFDs weighted by importance. But to say that moral deliberation itself is goal-oriented (or maximizing) is to misconstrue things—that’s just not what it looks like.

“Once we have established which features [of the circumstances] count which way, we make an attempt to decide where the balance lies. This is inescapably a matter for judgment, according to Ross, and theory cannot help at all. Theory could only help if we could rank our different prima facie duties... so that we knew in advance that, say, it is always more important to help others than to keep one’s promises. But no such ranking fits the facts” (Dancy, 1993, p. 221).

**Objection**: where do the PFDs come from? How are they justified?5

Reply: Ross thinks it suffices to appeal to their self-evident correctness:

“To me it seems as self-evident as anything could be, that to make a promise, for instance, is to create a moral claim on us in someone else... The main moral convictions of the plain man seem to me to be be, not opinions which it is for philosophy to prove or disprove, but knowledge from the start” (1994, p. 336).

For short, Ross embraces an **Intuitionist** epistemology of morality. In the next unit, we will see reasons to worry about appeals to intuition as justifications for how we might know some moral principle to be true or correct.

---

5For further discussion, see Davis (1993, pp. 211-3).
Chapter 4

Metaethics: A Quick Introduction

4.1 Taxonomy

Some things and distinctions it’s important to be clear about:

- truth/falsity
- language/mind/world (roughly: states of the mind either represent how the world is, or say how the world ought to be; language is a tool that people use to communicate their states of mind)
- volition/cognition, desire/belief (neutral terms: judgment, opinion)
- representational/non-representational states of mind and language
- properties/predicates
- speech-acts (assertion, commanding, interrogating)
- analytic/synthetic, a priori/a posteriori

**Realism** about morality (or, more strictly, moral discourse and practice) endorses:

- **Factualism + success.** Moral claims (‘Murder is wrong’) can be true or false (they are “truth-apt,” or suitable for truth/falsity). Some moral claims are, moreover, true. Contrast: imperatives, requests, hurrahs, questions.
• **Representationalism + success.** Moral judgments (being of the opinion that murder is wrong) are cognitive attitudes (a belief that murder is wrong). They aim to represent the world correctly (and some of them do!). Contrast: moral judgments express some kind of non-cognitive attitude ([dis]approval of murder).

• **Assertivism + success.** Moral claims are used to make assertions (statements that purport to be about matters of fact). Contrast: commands, requests, hurrahs.

Proponents include Michael Smith, Jonathan Dancy.

**Anti-Realism/Irrealism** denies one of these things. There are many kinds of anti-realism about morality. For instance.

**Error-Theory.** This view endorses:

• **Factualism – success.** Every moral claim is false (though truth-apt).

• **Representationalism – success.** Every moral belief is false (it aims to represent the world as being a certain way, it simply represents it incorrectly).

• **Assertivism – success.** Moral claims are used to make assertions (statements that purport to be about matters of fact). Contrast: commands, requests, hurrahs.

J. L. Mackie is the primary proponent.

**Non-cognitivism.** Each view in this class endorses:

• **Non-factualism.** Moral claims aren’t either true or false.

• **Non-representationalism.** Moral judgments are some kind of non-cognitive attitude (e.g., [dis]approval of murder).

• **Non-assertivism.** Moral claims are used to do something other than make assertions or try to influence the beliefs of our interlocutors.

To make things even more complicated, there are different kinds of non-cognitivism, and different kinds of nonassertivism. A very quick overview.
4.2. COGNITIVISM VERSUS NON-COGNITIVISM

- **Emotivism** (A. J. Ayer, “Ethics for Logical Positivists”). Moral judgments (e.g., the judgment that murder’s wrong) involve some kind of expression of emotion (*boo murdering!*). The speech-act performed by making a moral claim is something like exclamation (or, in Ayer’s unfortunate terminology, “ejaculation”).

  - This view needs to be distinguished (and Ayer does distinguish it, see pp.147-9) from a variety of realism known as speaker subjectivism (SS), according to which a speaker who says (judges) that *X is wrong* is asserting (believes) that she disapproves of *X*.

  - Discussion: Why is the SS view realist?

  - SS isn’t very plausible. Think of someone who says “if I didn’t disapprove of murder, it wouldn’t be wrong.” According to SS, they’ve said something trivially true. Most of us tend to think we’ve said something false.

- **Prescriptivism** (R. M. Hare). Moral judgments (e.g., the judgment that *A should X*) involve some kind of prescriptive attitude (wanting it to be the case that *A does X*). The speech-act performed by making a moral claim is something like prescription (recommending or advising that *A do X*).


- **Expressivism** (Simon Blackburn, Allan Gibbard). Moral judgments (e.g., the judgment that *A should X*) involve expressing acceptance of a system of norms (in this case, norms that require *A to do X*). The speech-act performed by making a moral claim is open: it could be just pure expression (like Ayer’s ejaculation), or perhaps an attempt to get you to coordinate your system of norms with mine.

4.2 Cognitivism versus Non-cognitivism

4.2.1 Why Be Non-cognitivist?

The Fact-Norm/Is-Ought Distinction

“Ethics, if it is anything, is supernatural and our words will only express facts; as a teacup will only hold a teacup full of water and if I were to pour out a gallon over it.”
Ludwig Wittgenstein (1994, #38) suggests that, if we collect or write down all the facts in some domain of inquiry, we will find “nothing which we could call an ethical proposition.”

Facts (and situations or states of affairs or propositions more generally) are one thing, norms are another. Given some state of affairs $S$, there is a bunch of stuff that is the case about $S$ (the facts about the events that are happening in $S$, for instance, it’s raining in $S$, I have an umbrella in $S$, I stay dry in $S$). But if we look at $S$ we never literally see anything like rightness or wrongness or goodness or badness.

Worries about this argument:

- Question-begging. If a moral claim is true, then, of course, it will appear in the collection of facts about the relevant state of affairs. Wittgenstein appears to be simply assuming that no moral claim could be a fact.

But it’s possible to strengthen the argument significantly:

1. **Sensory properties** (properties that involve the physical or observational properties of objects—properties that can be identified, either directly or indirectly, via sense perception) are the only properties that our scientific theories are able to recognize. (Assumption)

2. If a property $P$ is not recognizable by a scientific theory, then the claim $O$ has $P$ is not scientific. (Assumption)

3. Only scientific claims are factual (have truth-values). (Assumption)

4. Moral properties (rightness, wrongness, etc.) are not accessible to any faculty of sense perception (we don’t see rightness, nor do we hear it, etc.). (Assumption)

5. So, moral properties are not natural properties. (i, iv)

6. So, moral properties are not recognizable by scientific theories. (i, v)

7. So, the claim that an action has a moral property is not scientific. (ii, vi)

\[1\] There are good and bad reasons to support this premise. A good reason is just the idea that every fact is part of a true, possibly unknown, scientific theory about the world. A bad reason is Ayer’s **verificationism**—the idea that the factual content of a sentence is identified with the sensory experiences that would verify its truth (from which it follows that if there are no such sensory experiences, the sentence doesn’t have any factual content). No one thinks verificationism is true.
4.2. **COGNITIVISM VERSUS NON-COGNITIVISM**

8. So, the claim that an action has a moral property is not factual. (iii, vii)

Some available responses to this version of the argument:

- Deny that moral properties are not accessible to any faculty of sense perception. Two ways to go with this:

  - **Intuitionism/Non-Naturalism.** Moral properties are *non-natural*, have no causal powers, accessed by some distinctive cognitive moral faculty, some “special mode of cognition” (Dancy, 1993, p.411). The epistemology of moral facts\(^3\) is understood analogously to the epistemology of mathematical facts. (G. E. Moore, W. D. Ross, Jonathan Dancy)

  - **Naturalism.** Moral properties are natural—identical or reducible to scientifically respectable observational properties—in the same way that the property *being water* reduces to the property *being H\(_2\)O*. We come to know that things have moral properties in the familiar way; no special epistemology for moral facts required. (Peter Railton)

- Agree that moral properties are not sensory (not accessed by direct or indirect action of the senses), deny that this makes the claim that an action has a moral property non-scientific. Non-sensory properties can appear in scientific theories.

  - Take, for instance, one view of **secondary qualities**, like color. What our visual system perceives as a surface’s *redness* is actually, we know, a property about what wavelengths of (colorless) light are reflected by the surface. Redness isn’t actually hitting our eyes—it’s colorless reflected light. What we perceive as redness is the result of a dual contribution—part coming from the external world, part coming from our visual system, and being *projected* onto the world.

But we can still have a scientific theory of color, recognize color as a scientifically kosher property, and call color judgments factual. Perhaps moral properties are like secondary qualities. (John McDowell)

- It seems plausible that our experience of moral properties is like our experience of colors: the “raw material” of perception is processed by our brain *into* a morally laden perception. Wittgenstein even hints at

---

\(^2\)Here we just list them. We’ll come back to many of them in more detail a bit later on in the course.

\(^3\)I.e., our account of how people come to know that certain things have moral properties (are wrong/right).
something like this view: the philosopher confronts the paradox that “an experience, a fact, should seem to have supernatural value,” even though there is nothing intrinsic to the fact, scientifically considered, that appears to be this value (p. 145).

- There are worries for this sort of view: is the analogy between moral properties and color sound? And, even if it is, are there really facts about color, or is our visual system systematically misleading us—causing us to form beliefs about the colors of objects that just don’t match up with the world?

• Agree that moral properties are not sensory and that non-sensory properties cannot appear in scientific theories, deny that only scientific claims are factual.

  - This is available, but I don’t know how plausible it is.

The Argument from Queerness

“And similarly the absolute good, if it is a describable state of affairs, would be one which everybody, independent of his tastes and inclinations, would necessarily bring about or feel guilty for not bringing about. And I want to say that such a state of affairs is a Chimera. No state of affairs has, in itself, what I would like to call the coercive power of an absolute judge.” (Wittgenstein, p. 143)

The Argument from Queerness can be thought of supporting (i.e., as a response to any possible reply to) Wittgenstein’s non-cognitivist argument. The idea, in a nutshell, is that moral values don’t fit within a respectable scientific / naturalistic worldview. J. L. Mackie develops it best, and in two versions.

Version 1: Metaphysical
1. Moral judgments are categorical judgments—judgments of the form it’s wrong for A to X that attribute a property of categorical wrongness to A’s doing X (this means that the wrongness or normativity does not depend on A’s desires).

2. Normativity, insofar as we can understand it as being part of the natural world, is “constituted [and generally explained] by our choosing or deciding to think a certain way” (p.160). So far as we understand the natural world, we see there is no normative “pull” for A to do something (no motive for A) that does not involve an appeal to A’s prior choices or desires. If such normativity (“authoritative prescriptivity,” p.162) is real (exists), it exists outside of the natural world.
3. Nothing exists outside of the natural world. (Metaphysical naturalism)

4. So there is no property of categorical wrongness.

5. So no moral judgment is correct (Error Theory)

**Version 2: Epistemological**

1. If moral properties (objective values) exist, they exist outside of the natural world.

2. So, in order to be aware of moral properties, we would need some “special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else” (p.161) that brought us into contact with these properties. (In other words, intuitionism follows on the supposition that moral properties are exist and we are aware of them.) (i)

3. There are two kinds of faculty: sensory and non-sensory. (Assumption)

4. Our sensory faculties are not equipped to access such properties. (Assumption)

5. Our non-sensory (intuitive, a priori) faculties only furnish us with knowledge of analytical statements (and moral statements are not analytical). (Assumption)

6. So if we have knowledge about moral properties, that knowledge has to be synthetic a priori. (ii-v)

7. There is no synthetic a priori knowledge. (Assumption)

8. So, if moral properties exist, we can’t know anything about them.

**Motivational Internalism**

Michael Smith presents the argument from motivational internalism in (1994, #35):

1. Moral judgments necessarily do supply a motive (what he calls the “practicality” of moral judgment, and what’s known elsewhere as motivational internalism). Informally: judging that murder is wrong automatically gives you a motive/desire not to murder.

---

4The expressivist/emotivist will agree with the thrust of the argument, but deny that moral judgments involve representing some action as having some property at all.
2. But beliefs about matters of fact never supply a motive by themselves. (The facts are motivationally neutral—they don’t point us one way or the other.)

3. So moral judgments are different from beliefs about matters of fact—they must actually involve some motivational state.

“Psychologically, since making a moral judgment entails having a certain desire, and no recognition of a fact about the world could rationally compel us to have one desire rather than another, this seems to imply that our moral judgment must really simply be an expression of that desire.”

Note that the conclusion of this argument is non-cognitivism, although it has similarities to the Argument from Queerness: “The standard picture of human psychology tells us that there are no moral facts. Nothing could be everything a moral judgment purports to be.”

Smith ultimately rejects motivational internalism (terminology: he embraces motivational externalism), because internalism conflates moral judgment (judging there to be a moral reason for/against some action) with actually having a motive or desire (p.173). But Smith wants to separate these:

- To say that an agent A has a moral reason to X (that X is obligatory for A) is to say that A “would desire or prefer X if [A] was in “certain idealised conditions of reflection: if, say, [A] were well informed, cool, calm and collected” (p.174).
  - Why does Smith think this?
  - Why does this introduce a worry about relativism? (Hint: it has to do with his discussion of ideal convergence, pp.175-6.)
  - Why does Smith think it actually avoids the Argument from Queerness?

- So: judging that X is wrong = judging there’s a reason not to X = judging that I “would desire or prefer not-X if I was in “certain idealised conditions of reflection.”

- Although this rationally requires that we have the desire in our actual, non-idealized situation, we’re sometimes irrational—we don’t always desire what we’re rationally obligated to! The desire/motivation doesn’t always accompany the moral judgment.
4.2. COGNITIVISM VERSUS NON-COGNITIVISM

Dancy, on the other hand, rejects the argument while accepting motivational internalism (see 1993, p.415). According to Dancy, the facts aren’t always motivationally neutral—they’re not mere “inert states of the world.” Similarly, beliefs can sometimes supply a motive all by themselves.

- For his example of such a case, see p.416. We might wonder—this seems like an odd view about motivation. Certainly not the one that economists or artificial intelligence researchers use when they model decision-making.

Hume’s Law

Pidgen (p.423) describes the argument.

1. A moral conclusion can never be derived from exclusively non-moral (descriptive) premises. (A.K.A. no ought from is; Hume’s Law)

2. So, there is a basic distinction between moral and non-moral statements. (i)

3. All and only descriptive statements are factual. (Assumption)

4. So, no moral statements are factual. (ii, iii)

This is not a good argument.

- Premise (iii) just begs the question against the cognitivist.

- Moreover, as Pidgen notes, “we need not resort to non-cognitivism to explain this logical gulf. For there is a similar gap between conclusions about hedgehogs and premises which make no mention of them ... Yet nobody proposes a fact/hedgehog distinction or alleges that propositions about hedgehogs are not really propositions but a quaint subclass of commands” (pp.423-4).

4.2.2 Why Worry about Non-cognitivism?

Moral dispute. It seems like people can really disagree about moral questions. Normally we think that people disagree iff they contradict each other iff they assert contradictory propositions. If that is right, then non-cognitivism says that people never genuinely disagree about moral questions.

- Imagine that A hates ice cream, and B loves ice cream. A says ‘boo ice cream!’ B says ‘yay ice cream!’ Do A and B disagree with one another? Or do they just have different preferences/desires?
Ayer responds to this by denying (pp.150-1) that people ever really disagree about moral questions. People try to persuade others that they have the facts of a case wrong, and people often disagree about such facts. But once everyone agrees about the facts, and there is still a dispute, the argument ceases—at this point “we cannot bring forward any arguments to show that our system is superior” and tend to resort to verbal “abuse.”

Smith is less thrilled. People certainly behave as if there is a standard of truth against which to assess our moral positions. We are “preoccupied” with moral arguments; we expect people to supply reasons for and against their moral opinions; good arguments—where a moral conclusion is supported by good reasons—are likelier to “get the facts right” that bad ones. In other words, the norms and behaviors that govern moral arguments seem to be precisely the norms and behaviors that govern factual arguments.

Discussion: is Ayer or Smith right? Are there no moral disagreements that don’t reduce to arguments “about a question of logic or about an empirical matter of fact”?

Objectivity. If moral judgments and claims are factual, then there is a clear standard against which to measure their “validity,” and to assess moral arguments: the truth! But if they are not factual, there is no clear standard.

Ayer freely admits this: “We can now see why it is impossible to find a criterion for determining the validity of ethical judgments. It is not because they have an ‘absolute’ validity which is mysteriously independent of ordinary sense-experience, but because they have no objective validity whasoever” (p.148).

4.3 Naturalism versus Non-naturalism

Naturalism and non-naturalism are both varieties of realism. Naturalism says moral facts / properties reduce to natural ones (in the same way, e.g., water reduces to H$_2$O); what precisely they reduce to (pleasure, overall happiness, etc.) is a matter of controversy. Non-naturalism denies that moral facts / properties reduce to natural ones, although they usually agree that they supervene on natural ones.
4.3. NATURALISM VERSUS NON-NATURALISM

For $P$ to reduce to $Q$, there needs to be no real distinction (conceptual or otherwise) between $P$ and $Q$. Supervenience is weaker (reduction implies supervenience, but not vice versa): for $P$ to supervene on $Q$, it only needs to be the case that whether $Q$ holds is determined by whether $P$ holds.

4.3.1 Why be Non-naturalist?

Hume’s Law

A modified version of Pidgen’s argument.

1. A moral conclusion can never be derived from exclusively non-moral (descriptive) premises. (A.K.A. no ought from is; Hume’s Law)

2. If moral fact $M$ reduced to non-moral facts $N$, then from the premise that $N$ holds, we could validly infer that $M$ holds (since $N$ is identical to $M$).

3. So moral facts do not reduce to non-moral facts.

This argument proves too much. If $M$ supervenes on $N$, even the non-naturalist is committed to a version of (ii). So either Hume’s Law is false, or metaphysical relationships between kinds of facts (identity or supervenience) are not enough to license logical inferences.

- Plausibly it’s the latter. Most people agree that water reduces to H$_2$O, but I don’t think anyone would say that H$_2$O-facts, all by themselves, let you infer the same facts about water. You need bridge principles (what Pidgen calls “definitions”)—extra premises that bridge the logical/conceptual gap between H$_2$O and water (for instance, the premise that everything that is H$_2$O is water, and vice versa).

- This example shows that we can agree that there is a logical and conceptual gap/distinction between $N$ and $M$ (that might need to be bridged with bridge principles), without admitting an ontological distinction between them. Logical and conceptual gaps are compatible with reduction.

Pluralism

Dancy claims that pluralism makes non-naturalism look plausible (see 1993, pp.413-4).

1. Many different, apparently unrelated natural properties are intrinsically valuable. (Pluralism, assumption)
2. So, rightness (or goodness), if it is a natural property, is a **disjunctive property**: it’s gotten by taking all of the intrinsically valuable, apparently dissimilar natural properties \(P_1, \ldots, P_n\) and saying something is right (prima facie) iff it has \(P_1\) or \(P_2\) or ... or \(P_n\).

3. Natural properties are never disjunctive.

4. So rightness is not a natural property: “there is no common natural feature present for us to identify rightness with” (p.413).

I’m not sure I understand this argument. Are non-natural properties any more likely to be disjunctive than natural ones?

**Moore’s Open Question Argument**

For discussion, see especially Pidgen (1993, sect.iii) and Ridge (Stanford Encyclopedia). We let \(N\) be a natural fact, and \(M\) a moral fact, and suppose that \(M\) reduces to \(N\) (that’s to say: suppose that naturalism is right).

**Version 1.**

1. Since \(M\) reduces to \(N\), the question *does \(N\) hold?*, given that \(M\) holds, is unintelligible or trivial (like the question *does \(M\) hold?*, given that \(M\) holds).

2. This question is not trivial.

3. So naturalism is false.

Premise (i) is false. Water reduces to \(H_2O\), but the question *is this water*, on the supposition that it is \(H_2O\), is intelligible and not at all trivial. We need to know about the reduction—know about the lack of ontological distinction between two categories of properties/facts, know that the relevant bridge principles are true—in order for this sort of question to be trivial!

- In a way, our earlier discussion of Hume’s Law showed this: ontological reductions don’t require conceptual reductions. Two things can be associated with different concepts, and nevertheless be the very same thing.

- Examples of other *a posteriori* identities.
4.3. NATURALISM VERSUS NON-NATURALISM

Version 2.

1. Since $M$ reduces to $N$, the claim whenever $N$ holds, $M$ does too (e.g., the claim that whatever is pleasant is good) is trivial (like the claim whenever $N$ holds, $N$ holds).

2. But this claim is not trivial—in fact, it can, if true, give us a reason for promoting states of affairs in which $N$ holds!

3. So naturalism is false.

Premise (i) fails again, and for the very same reason. Triviality is a matter of conceptual content, not ontological distinctions (or lack thereof).

4.3.2 Why Worry about Non-naturalism?

Metaphysical Issues

The worry is the same as when Mackie voiced it: making moral properties non-natural only makes them more bizarre.

Dancy wants to resist this claim. We can say moral facts “exist in virtue of the non-moral ones.” Sometimes people say that moral facts supervene on (without reducing to) the non-moral or natural ones: the natural facts about a situation determine which moral duties there are in that situation, and if we had different natural facts, we might have a different duty in that situation. But the moral facts, although determined by the natural facts, are nevertheless different from the natural facts.

- Example (p.414): the fact that a cliff is dangerous depends on how steep and high up it is. But we wouldn’t want to say that the dangerousness of the cliff just is its steepness and highness. (For one, something can be dangerous without being either steep or high; dangerousness is multiply realizable.)

- This example is incredibly controversial. We might think that there is a natural feature that all dangerous situations have in common: being liable to harm an agent.

Epistemological Issues

Once again, the worry is Mackie’s: making moral properties non-natural only makes it harder to explain how we come to access/know the moral facts. We require some special cognitive faculty to enable us to access the moral facts, since
our ordinary sensory faculties are only capable of accessing natural facts.

Possible responses to this problem.

- Moral facts are known, like mathematics, via reason (Kant). So we do not require a special cognitive faculty to enable access to morality.
  This is worrying—is the knowledge that murder is wrong intuited in the same way that the knowledge that there are no round squares is intuited?

- There is a specialized, dedicated moral sense. Someone let the psychologists know!

- Praxis (Dancy, p.416): moral knowledge (like knowing how to do some task or achieve some goal) is a kind of praxis (practical knowledge or know-how). It’s about the agent reliably responding to reasons—the natural facts that determine whether or not some action should be done—in the right way. (Think of the difference between someone who succeeds in a complicated task by accident, and someone who succeeds by skill.)

  But if that’s right, then moral judgments might not be really cognitive at all! It’s very controversial in contemporary philosophy whether know-how (skill) can be reduced to ordinary knowledge: whether the knowledge we have when we know how to play guitar is just a matter of knowing a bunch of truths. For this view to be cognitivist, praxis needs to just be ordinary knowledge of a bunch of truths.
Chapter 5
Relativism, Objectivism, and Humeanism

5.1 Relativism

As Wong (1993, #39, p. 442) notes, relativism comes in two flavors. As he also notes, they can both be understood as reactions to the same phenomenon: deep, persistent, seemingly irresolvable moral conflicts.

- **Meta-ethical**: The truth of (or grounds for making) a moral assertion is somehow contingent or relative (depends on the culture, the speaker’s moral beliefs, the hearer’s moral beliefs, etc.). As Gowans puts it (§2):

  “[A] moral judgment such as ‘Polygamy is morally wrong’ may be true relative to one society, but false relative to another. It is not true, or false, simply speaking.”

Here are some versions of meta-ethical relativism, borrowed from Gowans.

- **Agent relativism.** A speaks correctly in making a moral claim that $X$ is wrong iff $A$ disapproves of $X$. (Slogan: the opinions of the speaker determine what’s right and wrong.)

- **Cultural relativism.** A speaks correctly in making a moral claim that $X$ is wrong iff $A$’s culture disapproves of $X$. (Slogan: the customs of the culture determine what’s right and wrong.)

- **Appraiser relativism.** A speaks correctly in making a moral claim that $X$ is wrong iff we, the people listening to (appraising) $A$’s claim, disapprove of $X$. (Slogan: the opinions of the listener determine what’s right and wrong.)
CHAPTER 5. RELATIVISM, OBJECTIVISM, AND HUMEANISM

The phrasing is designed to be neutral between Meta-ethical Realism and Anti-Realism.

- **Normative**: “It is wrong to pass judgment on others who have substantially different values, or to try to make them conform to one’s values.”

Note that normative relativism doesn’t necessarily take a stand about meta-ethical relativism. Indeed, you might embrace normative relativism because you endorse meta-ethical objectivism: you think it’s an important moral truth that the practices of certain cultures other than our own ought to be respected, for example. Normative relativism just requires a kind of “tolerance” (as Gowans puts it). But the value of tolerance might plausibly be an objective value!

We’ll focus our attention on meta-ethical relativism in this class.

### 5.1.1 Descriptive Moral Relativism

According to the argument from Descriptive Moral Relativism (associated with the character Protagoras in Plato’s *Theaetetus*; see Gowans, §2), we infer a certain kind of relativism just from the observation that a certain group $G$ (perhaps consisting of only the speaker) has different moral opinions than us.

1. We disagree with $G$ about some moral question $Q$.

2. There’s no reason to think that our opinion is any better than $G$’s. (No reason to privilege our opinion over $G$’s.)

3. So, neither of us is right. Our opinion about $Q$ is just as good as $G$’s opinion about $Q$.

Responses:

- People might disagree about premise (ii). We often evaluate people’s reasons for accepting certain opinions on moral questions. Some reasons are better than others. (For instance, the fact that you find abortion icky is not a good reason to oppose it.)

- Even if premise (ii) is true, the argument’s invalid. From the fact that we have no reason to privilege our opinion, it just doesn’t follow that there’s no fact of the matter about who’s right. There are lots of questions on which people disagree, where we have no good way of saying who’s right, but nevertheless we think there’s a fact of the matter. (E.g., is the number of blades of grass in Washington Square Park odd, or is it even?)

---

1As formulated by Gowans (see §2 of his essay), DMR says that “As a matter of empirical fact, there are deep and widespread moral disagreements across different societies.”
5.1. RELATIVISM

5.1.2 Functional Necessity

The argument from functional necessity (associated with Montaigne; see Wong, p. 443) notes that certain moral codes are essential for a group $G$ to function (or essential for the flourishing of people in $G$). What’s essential for $G$ to function might not be essential for a different group $G'$ to function. So what’s moral for $G$ might not be moral for $G'$. Gowans (§6) dubs this sort of view “pluralistic relativism”.

Responses:

- This isn’t relativism at all. It says that what’s right for an individual $A$ is a matter of what’s essential for the survival of the relevant group of which $A$’s a member. While that makes the correct moral code depend heavily on the characteristics of the group, it’s just ordinary situational variation; what best promotes some goal (the survival of the group) will vary depending on the circumstances. But consequentialism isn’t relativist because it acknowledges that what promotes the good will vary depending on the circumstances.

- The argument’s invalid. We might accept all of the premises, but still reject that what’s right for an individual $A$ is a matter of what’s essential for the survival of the relevant group of which $A$’s a member. Suppose $A$’s a skinhead. Do we want to say that what’s essential for the survival of the skinheads determines what it’s right for $A$ to do?

5.1.3 Persistent Disagreement

The argument from persistent disagreement trades on (i) the fact that when there’s an objective fact about some question $Q$, accumulating more and more evidence about $Q$ usually produces convergence about the right answer to $Q$; disagreement usually disappears; (ii) the fact that moral disagreement’s not like that—it’s deep and extremely persistent. Formally:

1. If there’s an objective fact about the matter about $Q$, then accumulating more and more evidence relevant to $Q$ will eventually produce convergence about the right answer to $Q$.

2. Much moral disagreement is deep and persistent (immune to the accumulation of evidence; note that this is just a consequence of DMR).

3. So, where disagreement about a moral question is persistent, there’s no objective fact of the matter about its correct answer.
To respond to this argument, we need some way of alleviating the discomfort (or, even better, some sort of explanation for why certain kinds of moral disagreement seem to be so persistent that’s consistent with there being objectively correct answers to moral questions). Some possibilities.

- Moral disagreement isn’t unlike disagreement about obviously factual questions when the stakes are high. When the stakes are high, people are just stubborn! The nature of moral disagreement is that the stakes are inevitably high.

- Persistent moral disagreement (e.g., between different cultures) can often be explained situationally: the two cultures may share some underlying value, but realizing that value in different circumstances requires different courses of action. (Wong’s example [p. 445]: supposing that societies all value reproduction, we would expect that polygamy’s acceptability would depend on the male:female ratio in a given society.)

### 5.1.4 Fundamental Diversity

The argument from fundamental, basic diversity (as opposed to garden-variety moral diversity) is the argument that Wong looks most favorably on (see p. 445). There are some moral differences that seem completely basic, and which can’t be explained by appeal to a single value that’s being implemented in different ways, because of different circumstances. For instance, certain moralities seem to place a fundamental value on family/kinship (on the duties an agent has in virtue of standing in familial relations to others), while others seem to place a fundamental value on the individual (on the rights an agent has just in virtue of being an agent).

> “Moralities, on this picture, are social creations that evolve to meet certain needs. The needs place conditions on what could be an adequate morality, and if human nature has a definite structure, one would expect further constraining conditions on an adequate morality to derive from our nature. But the complexity of our nature makes it possible for us to prize a variety of goods and to order them in different ways, and this opens the way for a substantial relativism to be true.” (Wong, p. 446)

More formally:

1. If there were a realm of objective moral facts, people would probably have equal access to them, and we’d expect some sort of convergence on basic questions of value. (The facts would decide whether kinship moralities
5.1. RELATIVISM

or rights-oriented moralities were correct and people would eventually converge on this.)

2. But there’s persistent divergence on basic questions of value.

3. The best explanation for this is that there’s not a realm of objective moral facts.

4. So, moral truth, if it exists, is not objective. It must be relative to something—a culture, group, society, or individual.

But Wong is careful to note that even if this sort of relativism is true, it doesn’t mean that any moral code is equally correct (that’s to say: it supports moderate rather than extreme meta-ethical relativism). There’s still a kind of objective basis for morality here—a morality has to meet certain basic social or human needs in order to be adequate. And certain moralities just won’t do a good job of this (e.g., moralities that “would aggravate interpersonal conflict”).

5.1.5 Internalism

There are two ways of running this argument. One’s better than the other.

The first: Recall Smith’s formulation of motivational internalism: “making a moral judgment entails having a certain desire” or motive. The argument for relativism from motivational internalism notes that this seems to imply that moral judgments require the presence of a certain desire in order to be true. So the truth of a moral judgment (made by an individual) depends on whether or not that individual has a certain desire. There are two problems with this way of running the argument.

• You might worry about motivational internalism. (Smith rejects it!)

• Motivational internalism doesn’t imply that moral judgments require the presence of a certain desire in order to be true. It says that if the agent didn’t have the desire, she wouldn’t have made the moral judgment. But that has nothing to do with the moral judgment’s truth or correctness. The conditions needed for an agent to make a moral judgment ≠ the conditions under which that judgment is true.

The second: the argument from Internalism about Moral Reasons.

---

2 This way is originally due to Gilbert Harman, “Moral Relativism Defended” (The Philosophical Review 84 [1975]; available at http://www.jstor.org/pss/2184078).
1. The truth of a moral claim (that A should X) implies that A has a reason to accept that moral claim. ([Internalism about Moral Reasons]: morals give reasons)

2. A has a reason to accept that she should X only if she has a desire that is satisfied by X. ([Humean Theory of Rationality/Reasons])

3. So the truth of a moral claim addressed to any agent depends on whether that agent has a certain desire.

The worry with this argument is with the Humean Theory of Rationality, which Smith also rejects. We’ll discuss it in depth below.

5.2 Objectivism

As defined by Gowans (§2), Objectivism is the view that:

“[M]oral judgments are ordinarily true or false in an absolute or universal sense, some of them are true, and people sometimes are justified in accepting true moral judgments (and rejecting false ones) on the basis of evidence available to any reasonable and well-informed person.”

In a phrase, Objectivism claims that moral truth doesn’t need to be qualified (as truth for some culture or individual). It’s ordinary, honest-to-goodness truth for everyone.

Let’s look at some arguments for objectivism (or, equally, against meta-ethical relativism).

5.2.1 Relativism Is Contradictory

This argument (see Gowans, §5) goes like this:

1. If meta-ethical relativism is right, then murder is wrong (since this is true for some group) and murder is not wrong (since this is true for a different group).

2. It’s not possible that murder both is and is not wrong.

3. So meta-ethical relativism is false.

The relativist response:
5.2. OBJECTIVISM

“The standard relativist response is to say that moral truth is relative in some sense. On this view, S is not true or false absolutely speaking, but it may be true-relative-to-X and false-relative-to-Y (where X and Y refer to different societies or moral frameworks, and their respective standards). This means that suicide is right for persons in X, but it is not right for persons in Y; and, the relativist may contend, there is no inconsistency in this conjunction properly understood.” (Gowans, §5)

We only have a contradiction if murder is both wrong and not wrong relative to X. But we don’t.

5.2.2 Relativism Makes Moral Disagreement Impossible

Suppose A says ‘murder is wrong,’ speaking from one moral framework (according to which murder is wrong), and B says ‘murder isn’t wrong,’ speaking from a different moral framework (according to which murder’s ok). They’re both right (they both speak truly or correctly)! So there’s nothing to fight about, and they don’t actually disagree.

This is a strong worry, and it’s not clear what the relativist has to say about it. But note that pretty much everyone has this worry (difficulty explaining what’s going on when people have moral disagreements), except the realist. (See, again, Gowans, §5.)

5.2.3 Relativism Is No Better than Objectivism

“The specification of the relevant group is itself a morally significant question, and there appears to be no objective map of the world that displays its division into social groups to which the truth or justification of moral judgments are relative” (ibid.)

The relativist says truth is relative to the authoritative standards of, say, a society. As Gowans notes, “an explanation is needed of why some standards are authoritative for people in a society” (ibid.). But there’s disagreement within societies about moral questions and standards: what makes one social faction’s conception of the standards more authoritative than another’s? The fact that it’s held by a majority?

These seem like hard questions to answer. Moreover, compare them to the objectivist’s task: she needs to say which standards are authoritative for everyone, period. And that doesn’t actually seem much harder than saying which standards are authoritative for an entire society (unless there’s perfect agreement in
that society). The objectivist faces precisely the same difficulties that the relativist faces.

There are four reactions the relativists could have here:

- Affirm that, “when a dissident challenges some of the fundamental standards of his or her society... this person [is] necessarily wrong” (ibid.). Good luck with that!

- Stipulate that the groups that moral truth is relative to never have any disagreement about the standards. The authoritative standards are just the standard of the group. (This has the effect of making the groups that moral truth’s relative to quite small, perhaps consisting of just one person, or extraordinarily homogenous.)

- Decide that objectivism is no worse than relativism, so go back to objectivism.

- Decide that objectivism and relativism are equally bad, so ditch both for some form of anti-realism.

5.3 Humeanism

Thomas Nagel (1994, #41) asks us to consider the example of someone who decides to steal a book (let’s suppose it’s rare) from the library. We reply to him: you shouldn’t do that; it’s unfair to others who might wish to borrow the book. He replies: I don’t care. Nagel’s reaction:

“The fact that he doesn’t care doesn’t make it all right: he should care. But why should he care?” (ibid., p. 156)

There are various (and distinct!) questions we can ask about our friend:

- Does he have a motive to care about others?
- Is he being irrational in failing to care about others?
- Is he ignoring a reason that he has to care about others?
- Is he morally in the wrong in failing to care about others?

For our purposes, the second and third question will amount to the same thing: being rational and having a reason are the very same thing.\(^3\)

\(^3\)Formally: A is rational to X iff A is reasonable to X iff A has a reason R to X and there’s no reason R’ > R to not-X.
Hume’s view, suitably understood, can be seen as giving answers to each of these questions.

5.3.1 Motivation

“Reason alone can never be a motive to any action of the will [and] can never oppose passion in the direction of the will.” (Hume, 1994, p. 118)

Motives are understood as the psychological causes of an action (whatever psychologically explains why an agent did some action).

According to the Humean Theory of Motivation, $A$ is motivated to $X$ only if $A$ wants $Y$ and $A$ believes that doing $X$ will get her $Y$. Beliefs alone (Hume’s “Reason”, a psychological faculty, not to be confused with having a reason) can never motivate.

So what is the motivational role of belief?

“Abstract or demonstrative reasoning ... never influences any of our actions, but only as it directs our judgment concerning causes and effects ... The impulse arises not from reason, but is only directed by it.” (ibid.)

Beliefs give instructions about how to satisfy our desires. Desires set the agenda; the role of beliefs is to find out how to implement the agenda. In Hume’s piquant formulation, Reason is “the slave of the passions.”

Why think this is true? Mainly, we look to examples of when agents are actually motivated. According to Hume, we discover that:

“It can never in the least concern us to know, that such objects are causes, and such others effects, if both the causes and effects be indifferent to us. Where the objects themselves do not affect us, [knowledge of] their [causal] connexion can never give them any influence; and ‘tis plain, that as reason is nothing but the discovery of this connexion, it cannot be by its means that the objects are able to affect us.” (ibid.)

5.3.2 Rationality and Reasons

“[A]s nothing can be contrary to truth or reason, except what has a reference to it, and as the judgments of our understanding only have this reference, it must follow, that passions can be contrary to reason only so far as they are accompany’d with some judgment or opinion.” (p. 120)
The argument here is this. First, we have to distinguish between basic and non-basic desires. A’s desire for \( X \) is \textbf{non-basic} only when A has some other desire \( Y \) such that A desires \( Y \) and A believes \( X \) will get her \( Y \). A’s desire for \( X \) is \textbf{basic} only when it is not non-basic.

1. A mental state can be rationally criticized (as false or contrary to reason) only if it represents or relies on the world (wrongly) as being a certain way.

2. Agents’ desires rely the world (wrongly) as being a certain way only when they’re non-basic and founded on a false cause-effect belief (that satisfying the desire will get her something else she wants).

3. So, basic desires cannot be criticized.

The \textbf{Humean Theory of Reasons/Rationality}: “‘Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger.” (p. 120)

According to Hume, why is it that people tend to think his theories are false (see the discussion of calm and tranquil passions—passions that are experienced without strong emotion, p. 121)? Why are these not good worries?

\textbf{Worry One}. If Smith is right, A has a reason to \( X \) iff A would want/prefer/care about \( X \), if she were made ideal (fully informed about the relevant facts, cool, calm, collected). It can be true that A would want/prefer/care about \( X \), if she were made ideal, even though A, as she is now, doesn’t currently want/prefer/care about \( X \). So if Smith’s right about reasons, Hume’s wrong: you can have a reason to \( X \) without currently wanting/preferring/caring about \( X \).

\textbf{Worry Two}. Nagel, on the other hand, thinks that reasons can be a matter of consistency. Consider your friend again. Suppose you ask your friend how he’d like it if someone else stole the book from the library, so that he’d never have a chance to borrow it. Suppose he says he wouldn’t like it—indeed, he’d \textit{resent} it. Why does that matter? This depends on what it means for someone to resent someone else’s actions.

1. \( A \) resents \( B \)’s action \( X \) in situation \( S \) only if \( A \) thinks \( B \) shouldn’t have done \( X \) in \( S \).

2. \( A \) thinks \( B \) shouldn’t have done \( X \) in \( S \) only if \( A \) thinks \( B \) had a reason not to do \( X \) in \( S \).
3. **Reasons depend only on general features of the situation** *(see p. 158).* $B$ has a reason not to do $X$ in $S$ iff anyone in a situation qualitatively similar$^4$ to $S$ has a reason not to do $X$.

4. So, if $A$ thinks $B$ had a reason not to do $X$ in $S$, she is committed to thinking that anyone in a qualitatively similar situation to $S$ has a reason not to do $X$.

5. So, if $A$ thinks $B$ had a reason not to do $X$ in $S$, she is committed to thinking that if she were in a qualitatively similar situation to $S$, she’d have a reason not to do $X$.

If Nagel’s right, Hume may be in trouble. If someone thinks that desires supply an agent’s reasons in a situation, she is committed to thinking that any agent in a qualitatively similar situation (with the same desires) would have the very same reasons. And agents just don’t think that any agent in a qualitatively similar situation (with the same desires) would have the very same reasons, as the phenomenon of resentment (when another person fails to respect their interests and concerns) shows.

Three available responses to Nagel’s argument.

- Doesn’t get any purchase on Hume, since Hume would not endorse the reaction of resentment (notice what he says about the person who prefers the destruction of the world to her suffering some minor inconvenience).

- Doesn’t get any purchase on agents who wouldn’t resent the thing they’re doing to others. (But such cases are very rare indeed, Nagel thinks. Most people take “their own interests and harms [to] matter, not only to themselves, but in a way that gives other people a reason to care about them too” [p. 158].)

- He may be wrong about reasons: agents can consistently think that reasons depend on agent-relative features of the situation (features of the situation that involve them essentially). $B$’s reason not to do $X$ in $S$ is that it harmed me! I’m just that important! (Perhaps, though, very few people would ever think such a thing. And perhaps that diminishes the force of the objection.)

$^4$Qualitative features are just general features of the situation: what general sorts of actions are available, what consequences those actions will have, etc. None of these things mention the identities of anyone that’s part of the situation.