

A decorative graphic consisting of a thin gold circle on the left side. A large black bracket is positioned on the left, with its top and bottom arms extending upwards and downwards respectively, and its right arm pointing towards the center. A large gold bracket is positioned on the right, with its top and bottom arms extending upwards and downwards respectively, and its left arm pointing towards the center. The text 'Value Theory' is centered within the space between these brackets.

Value Theory

PHIL 640

Preliminary matters

- Fill out cards with basic info for my records. Ignore items intended for ug's, and instead just list/describe prior work in ethics or other relevant areas, along with an indication of your own central interests.
- Can someone (or several people) volunteer to fetch the box of projector wires from the office before class and return it afterward? I also need a tall person to turn the projector on and off.
- And while I'm at it, who will remind us to break halfway through? And could someone help me schlep stuff back to my car?

[Readings]

- **Joyce, *The Myth of Morality*: an argument against moral truth, to be read critically and used as a way of organizing classic readings on metaethics and practical rationality.**
- **Darwall, Gibbard, Railton (eds.), *Moral Discourse and Practice [DGR]* : a well-known anthology bringing together some of the work that Joyce discusses,**
- **Also, some supplementary readings on my website, e.g. recommended background readings from Hume's *Treatise* and Kant's *Groundwork*. To access these, use "Greenspan" (case-sensitive) as both your username and your password.**

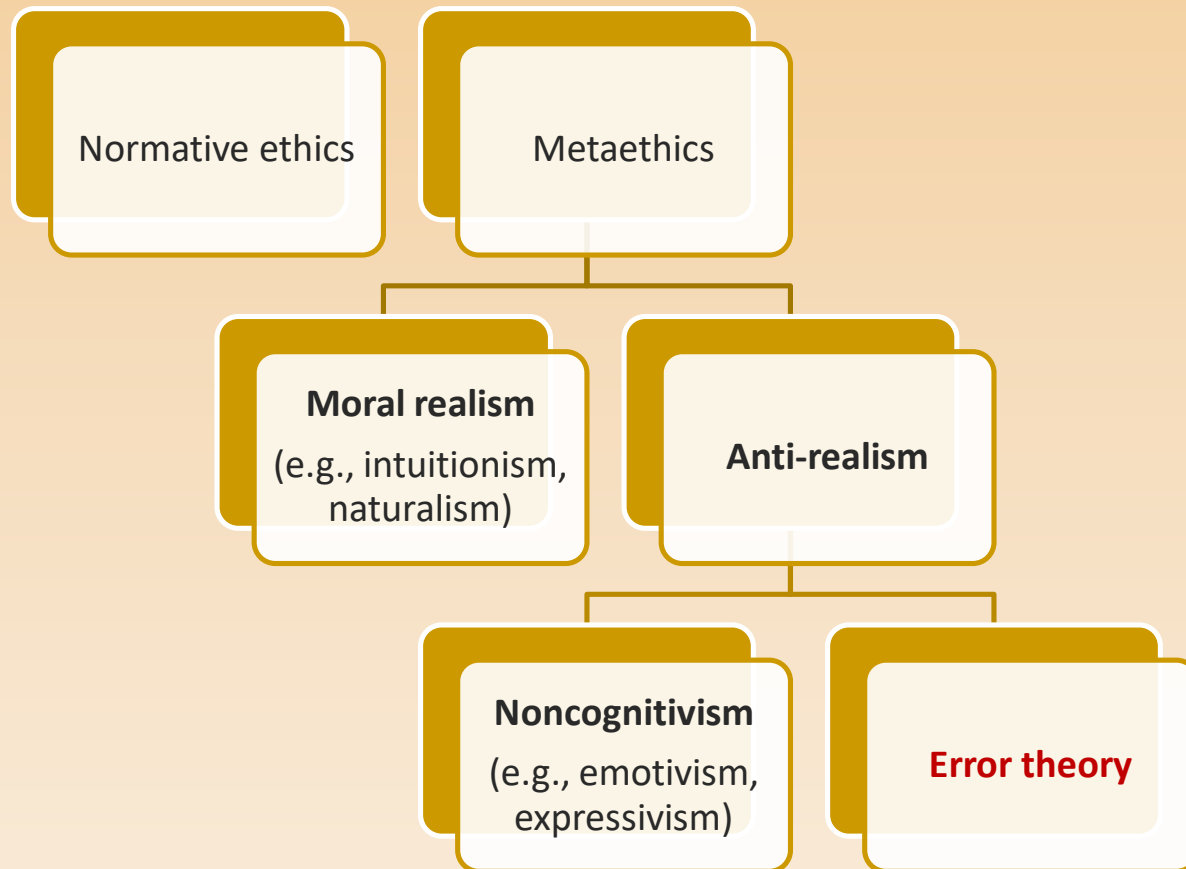
Requirements

- Midterm, final exams (take-home; 1-2 essay questions).
- Presentations on some of our later readings (basically just leading discussion, ideally with someone else)
- Possible one-page expositions and/or criticisms of the weekly readings
- Only the exams will be assigned letter grades. I use your ungraded written work and oral participation as a way of possibly boosting grades from what you get on the exams (up to one full grade).

Initial assignments

- We'll start slowly, with Joyce's Preface. I'll initially focus on providing some historical context, so I've posted selections from Hume's *Treatise* as recommended background reading. You might also (or instead) read Moore in DGR as background to Stevenson, though Joyce doesn't bring him up till later.
- For the following week I've assigned Joyce's discussion of noncognitivist anti-realism, along with the DGR excerpt from Stevenson as a sample. But Stevenson's reading of Hume needs serious correction, and some of his assumptions are taken from Moore.
- Next we'll read Mackie on error theory, along with Joyce's version of his view in the second half of ch. 1.

Some basic contrasts



[Humean background]

- Hume is the first pre-20th-century figure who focused distinctly on metaethics as the basis for his ethical theory, though his place on the chart is disputed.
 - Many noncognitivists represent him as an early proponent of their view – though, oddly, Stevenson interprets him as holding (“in effect”) that moral judgments just describe majority approval, not necessarily the speaker’s.
 - On the other hand, Mackie claims Hume as a forerunner of error theory, the cognitivist anti-realist view that Joyce will be defending.
- Some contemporary philosophers question whether Hume is really such a skeptic about morality. He sees moral properties as not really located in objects, but rather in ourselves, by analogy to Lockean secondary qualities, but they’re still independent of individual minds.

Hume's metaethical starting points

- Hume denies that reason can be the source of moral judgment. Its role is limited to discovering truth or falsity.
- Morality motivates us, and for motivation we need a passion (broadly construed to include dispositions such as benevolence, as “calm passions”).
- Neither a passion nor an action can “represent” anything – or be said to be true or false, reasonable or unreasonable, except derivatively, on the basis of accompanying factual judgments.
- So there's a gap between “is” and “ought” – which Hume thinks we need to fill in with facts about our moral sentiments.

Hume's moral sentiments

- Hume gives a two-stage account of the development of moral sentiments from passions (emotions and emotional dispositions) assumed to be built into human nature.
- He sees moral sentiments as primarily directed toward agents and their characters and motives, only derivatively toward acts. So in normative terms he's a virtue ethicist.
- At the same time, though, he's a forerunner of utilitarianism, since he rests the virtues on utility (construed in terms of pleasure/pain).
- However, unlike a full-blown utilitarian he has in mind utility for either the agent himself or others the agent is likely to interact with – as opposed to all mankind, or all sentient beings.

Sympathetic passions

- The first stage of Hume's account turns on sympathy with others harmed or benefited by an agent.
- The initial basis for a moral sentiment simply amounts to the pleasure/pain we feel at benefit/harm to ourselves . By associating these impressions with an idea of the agent who gives rise to them, we form passions of love or hate, pride or humility.
- These and other feelings then get extended, via sympathy, to cases where others are benefited/harmed by someone.
- However, since sympathetic passions vary with our standpoint of observation – we react more to agents whose behavior affects others closer to ourselves – we don't yet have moral sentiments.

Correcting sympathetic passions

- A second stage, of correction, is needed to allow for consistent uses of moral language across different standpoints of observation, as required by the communicative use of moral judgment.
- The corrected passion reflects a “general” or “common” point of view, the standpoint of an ideal (= ideally situated) observer or spectator, understood as someone in the circle of those affected by the agent’s actions.
- It’s only after this correction that we can be said to have moral sentiments, so they needn’t reflect our immediate reactions (“gut feelings”). Sometimes they even rest solely on supplying a term, without any corresponding feeling, for the sake of consistency with similar cases.

[The Kantian Alternative]

- When twentieth-century social science made people more aware of cultural differences, Hume's reliance on a universal "human nature" came to seem problematic, and many philosophers treated Kant's approach as the alternative.
- Kant offers an ethics of duty that doesn't depend on contingent human nature, but instead is supposed to hold necessarily for all rational beings who don't necessarily act in accordance with reason.
- His central concept is that of a categorical imperative = an unconditional "ought" that gives reasons to anyone, regardless of his desires.
- Even skeptics like (early) Foot and Williams understand morality in Kantian terms, as we'll see, and Joyce's argument follows suit.

Twentieth-century metaethical starting points

- Many current non-specialists in ethics (e.g. writing about evolutionary ethics) conflate Hume's "is"/"ought" gap with the "naturalistic fallacy," derived from 20th-century moral philosopher G.E. Moore.
- But what Moore had in mind concerned *definition*: an inference from a general characterization of the properties basic to ethics – he focused on "good" -- to a definition in terms of naturalistic or other factual properties (e.g., pleasure, as suggested by Mill).
- Moore comes up in Joyce much later, but he essentially kicked off 20th-century metaethics with a "nonnaturalist" version of moral realism, claiming that "good" refers to an indefinable property of objects, known by a faculty of intuition.

Moore's "open question" argument

Moore's central argument for nonnaturalism is taken for granted by Stevenson and most other 20th-century authors, so here's a sketch:

1. For any natural property N, suppose good =_{df} N.
2. Then it shouldn't even make sense (= be an "open question") to ask whether something you recognize as N is good. It would be tantamount to questioning a tautology, asking: "Is this good thing good?"
3. But for any N, the question whether an N is good makes sense (is intelligible, or "open," even if we think the answer is "yes").
4. So for any natural property N, good ≠_{df} N.

The move to anti-realism

- Ross and others followed Moore in defending a view known as “intuitionism,” extending to “ought” as well as “good” (whereas Moore initially had held a version of utilitarianism, with terms for right action defined in terms of “good”).
- Besides its mysterious epistemological basis, intuitionism faced problems in explaining widespread moral disagreement – particularly with increasing awareness of other cultures.
- At the same time, general philosophic authors who embraced logical positivism, such as A. J. Ayer, could find no room in their epistemologies for moral knowledge. Later, linguists began to emphasize other possible functions of language besides describing putative facts (assertion).

[Joyce's project]

- Joyce has two main aims, as outlined in his Preface:
 1. Establish *error theory* (that moral discourse, taken as requiring categorical force, is defective, since it's assertoric but fails to state truths)
 2. Defend *fictionalism* (substituting nonassertoric uses of moral language, as a kind of group pretense, to retain the motivational benefits of morality)
- [Many have questioned whether 2 would work, and I'll argue that 1 rests essentially on combining a Kantian interpretation of morality with a Humean understanding of reasons (as in Williams).]

[The error]

- Joyce wants to argue that moral discourse is analogous to 16th-17th century (pre-Lavoisier) phlogiston discourse, with phlogiston seen as an invisible substance released during combustion (vs. oxygen, which is taken in).
- The analogue for moral discourse would be claims about what we “must” do – without reference to institutional rules, and whether or not it promotes our aims or self-interest.
- Redefining “phlogiston” or strong moral “ought” (= “must”) in a way that would escape the error would deprive the relevant discourse of its point.

Carrying on anyway

- Unlike phlogiston discourse, though, moral “must” has evolutionary sources, and doing without it would carry a practical cost. But so would commitment to untrue beliefs.
- Joyce therefore proposes that we accept morality as a “fiction” and continue using it – like a “myth,” understood as a false narrative important to a group and therefore “accepted” by them (in a sense weaker than belief).
- [Ultimately, his argument will depend on the tendency of even unasserted “must”-thoughts to evoke emotion. But among other things we’ll want to ask whether this depends on belief at an earlier stage, in childhood.]

Stevenson's emotivism

- In a view that allows for descriptive as well as emotive meaning, Stevenson provides a sophisticated version of emotivism, the view outlined by Ayer in primitive (“Boo/Hurrah!”) form.
- Whereas Ayer had interpreted ethical statements simply as descriptively meaningless expressions of the speaker’s emotions, Stevenson assigns them a further, emotive *form* of meaning that doesn’t depend on describing distinctively moral facts.
- Instead of describing a natural *or* non-natural property of objects, “good” adds emotive meaning to the description of the speaker’s “interests” (= emotions and other pro/con-attitudes) – thereby leaving room for Moore’s “open question.”

Requirements for a “relevant” definition of “good”

1. makes sense of disagreement

- vs. Hobbes’ subjectivist “interest” theory, which makes disagreement impossible, since we’re each just reporting our own desires
- Stevenson substitutes “disagreement in interest” = opposing inclinations

2. possesses a certain “magnetism” (= motivational force; cf. p. 73)

- vs. Hume’s [alleged] interest theory reporting majority approval
- cf. contemporary [motivation] internalism [and Hume’s *Treatise* III.I.II]

3. isn’t confirmable solely through scientific method

- vs. all previous (i.e. purely descriptivist) interest theories
- rests on accepting Moore’s “open question” argument against naturalism

From dynamic use to emotive meaning

- Utterances with descriptive meaning can sometimes be used “dynamically,” to get people to do or feel something -- e.g. “It’s hot in here,” as used to get someone to open the window.
- Emotive meaning amounts to the *tendency* a word has, by virtue of its history of dynamic use, to produce and/or result from affective responses.
- On Stevenson’s account a statement like “x is good” does describe our interests, but also, and more importantly, it *expresses* them, in a way designed to subtly influence others to share them (e.g. via emotional contagion).

[Two approximations]

- Stevenson offers two attempts at defining “x is (non-morally) good,” though he concludes that neither is adequate:
 - “We like x” [used dynamically].
 - “I like x; do so as well.”
- These give a rough description of what we mean, but not a definition, since they lack the requisite subtle influence and hence can’t serve as “relevant” [i.e., adequate?] substitutes for “good.”
- Stevenson ultimately interprets the question whether something is good as a *request for influence* – in the first instance, by citing reasons, but once we agree about the facts, by non-rational means of persuasion.

Some standard objections

- It certainly looks as though we *think* we're describing something beyond our own reactions when we use moral terms.
- Also, Stevenson's view limits the role of rational argument in ethics to clearing up disagreement about nonmoral facts, but can't we argue more or less rationally about value? A view that doesn't allow for that seems unsatisfying to many philosophers.
- A more specific worry about rational moral argument is known as the Frege-Geach problem: whether and how we can make univocal sense of moral language in non-assertoric contexts (e.g., "if"-clauses, negation), of the sort essential to logical argument.

Joyce on defective discourses

- Joyce holds that moral discourse amounts to a “faulty framework,” of the sort exemplified by defective concepts that are based on “non-negotiable” untrue propositions, such as
 - *phlogiston* (stuff stored in bodies and released in combustion, primary component of soot), and
 - *tapu* (contagious impurity).
- Non-negotiability is determined by a “translation test” (pp. 4f.) that Joyce will use in his own ensuing argument: asking whether users of the discourse would be willing to take the term in question as translating a term in another language that *isn't* held to satisfy the untrue propositions.

Error theories

- An “error theory” holds that statements in a certain discourse are used to make assertions but typically fail to state truths.
 - An argument for an error theory has two parts: *conceptual* (identifying non-negotiable elements of the meaning of a term) and *ontological/substantive* (determining that nothing satisfies them).
 - An error theory is often thought to maintain that statements in the discourse are typically *false*, but Joyce adopts an interpretation of them as neither true nor false to accommodate Strawson’s account of non-referring expressions.
- Noncognitivism, the contrasting anti-realist view, is often characterized as denying that moral statements have truth values, but Joyce reformulates it as holding that moral judgments are typically used, not to assert something, but rather to express approval or a disguised command.

Noncognitivism

- Error theory clearly fits phlogiston discourse, but is noncognitivism more plausible for *tapu* (and moral) discourse? Not unless we take it as meaning something other than users of the discourse seem to think it means – as spelled out by Glassen’s earmarks of assertoric discourse; see p. 13).
- Joyce treats Stevenson’s version of noncognitivism as *pragmatic*, in contrast to Ayer’s less plausible semantic version: the primary *use*, rather than the meaning, of moral language is non-assertoric. [But cf. Stevenson’s “emotive *meaning*.”]
- He thinks that the shift from meaning to use lets Stevenson avoid the Frege/Geach problem (see pp. 10-12), since moral judgments retain their meaning [??].
- Instead, he objects that we need to retain the appearance of assertion in order to manipulate others, by seeming to provide them with a reason. So on a Stevensonian account the influence of moral statements would seem to depend on either belief or feigning belief (see pp. 14f.), i.e. asserting something objective.

[Fair against noncognitivism?]

[Does Stevensonian magnetism really *require* (feigning) belief?

- Can't Stevenson rely on persuasion by emotional contagion and other forms of nonverbal, or at least nonassertoric, social influence? Cf. also expression of *second-order* emotions (as in Blackburn's "emotional ascent"), e.g. to shame an opponent of "the dole."
- Cf. Gibbard's *expressivism*, which understands moral utterances as expressing a psychological state of *norm-acceptance* – rather than either belief or emotion (though the norm in question is a norm for assessing certain emotions). We move toward consensus through noncognitive influence and pressure toward consistency in group discussion, with the function of promoting social coordination.]

Mackie's cognitivist alternative

- Mackie accepts noncognitivism as a *revision* of our ordinary moral judgments, applicable to his own claims about normative ethics in the rest of the book from which the DGR excerpt is taken.
- But on his account ordinary moral judgments involve a false belief in objective moral values.
- He initially describes his metaethical position as a second-order version of skepticism, or a variant of subjectivism. But because of the confusing array of distinct positions coming under those terms, it's generally referred to as (an) "error theory" (cf. p. 94).

Shifts in focus

- Note that Mackie, in contrast to Stevenson, focuses primarily on characterizing moral requirements, rather than defining “good.”
- Also, Moore’s nonnaturalism and Rossian intuitionism had fallen out of favor by the time he wrote, so Mackie’s characterization of recent swings between metaethical views is limited to noncognitivism and naturalism. What had intervened were
 - a noncognitivist alternative to emotivism with Kantian elements: Hare’s “universal prescriptivism” (with moral judgments taken as expressing commands applicable to all relevantly similar agents) and
 - variants of utilitarianism and virtue ethics, with a basis in facts about human nature.

Kantian assumptions

- Mackie rejects both noncognitivism and naturalism as incapable of capturing the *authority* of ethics, understood in terms of the Kantian notion of categorical imperatives.
- In Kant's terms, moral requirements are necessarily binding, or "inescapable," applying to all rational beings, whatever their ends.
- What Mackie finds objectionable is the "objective, intrinsic prescriptivity" of moral requirements: their force for action, combined with their independence of our actual motives.
- [Nonmoral good and similar evaluative concepts lack this aspect and hence we can count them as objective without a problem.]

Mackie's arguments

- ***from relativity***: Cultural variation is better explained as reflecting divergent ways of life than in terms of something objective viewed from limited perceptual/evidential standpoints.
- ***from queerness***: The relevant property, “objective prescriptivity,” would be
 - bizarre [or even incoherent?],
 - inexplicably related to natural properties, and
 - unknowable, except through a mysterious faculty of moral intuition.
- ***from objectification***: Talk of objective moral properties can be better explained as a projection of social demands onto the world in order to make them seem authoritative.

Source of the “queerness”?

- As Mackie states it, his central argument, from queerness, seems to rest on combining objectivity with something like Stevensonian “magnetism” = motivational force.
- As with Moore’s “open question” argument, even later authors holding contrasting positions had tried to accommodate the [motivation(al)] internalist element of noncognitivism, i.e. motivational force as internal to the meaning of a moral judgment. Mackie sometimes seems to be claiming that this rules out mind-independence.
 - However, Joyce will interpret “prescriptivity” as normative rather than motivational. Mackie’s terms “to-be-doneness” (for acts deemed right) and “to-be-pursuedness” (for ends deemed good) involve *reason-giving* force.
 - As Joyce acknowledges, though, at least part of what Mackie himself represents as bizarre is the thought that something independent of our motives could *make* us act, in the manner of Plato’s Form of the Good (see DGR, pp. 96f.).

Ethics without objectivity?

- Mackie apparently thinks we can just go on and do ethics (at any rate, utilitarian ethics) after rejecting objective moral values – since he has no problem with nonmoral values such as rationality as applied to social ends.
- But could this sort of noncognitivist revision of ethics have the same kind of influence on others, unless it relies on their tendency to misinterpret it as objectively prescriptive?
- Joyce’s “fictionalism” will attempt to provide an alternative: a way of securing the social benefits of belief in objectivity, though we realize it’s just a pretense and aren’t relying on others’ being misled.

[Joyce on Mackie]

- Joyce divides Mackie's argument for error theory into two basic steps that he'll retain:
 - Conceptual: Moral discourse is committed to thesis T.
 - Ontological/substantive: T is false.
- For Mackie, T amounts to the claim that moral properties are “objectively prescriptive”: the universe issues prescriptions, in Joyce's terms.
- Joyce goes on to examine two ways of understanding T's reference to prescriptivity: as a thesis about motivation (in the remainder of ch. 1), which he ultimately rejects, and as a thesis about the *reasons* provided by moral discourse (ch. 2).

Motivation internalism?

- Brink's argument against Mackie depends on interpreting Mackie's T as ascribing (defeasible) motivation to anyone who sincerely holds a moral judgment; see MI, p. 18 [= motivation internalism, aka judgment-internalism].
- Although objections to MI have centered on the possibility of a totally amoral [= morally unmotivated] agent, the moral claims of such an agent could be said just to be reports of others' opinions.
- Joyce instead argues for the possibility of a "purely evil" agent whose actions are *explained* by her sincere moral judgments, even in opposition to others' opinions.

“Pure evil”

- Joyce’s example is Sade’s Eugenie, who takes her moral judgments (sometimes in disagreement with others’ judgments) as reasons to do what they forbid.
- Eugenie isn’t weak-willed, so she’s also a counterexample to Smith’s reformulation of MI as MI* (p. 22), which assumes rationality on the part of the agent.
- However, it’s not so easy to show that moral discourse is non-negotiably committed to MI/MI*, as indicated by uncertainty about how we’d translate aliens’ nonmotivational use of “schmood” (pp. 26f.).

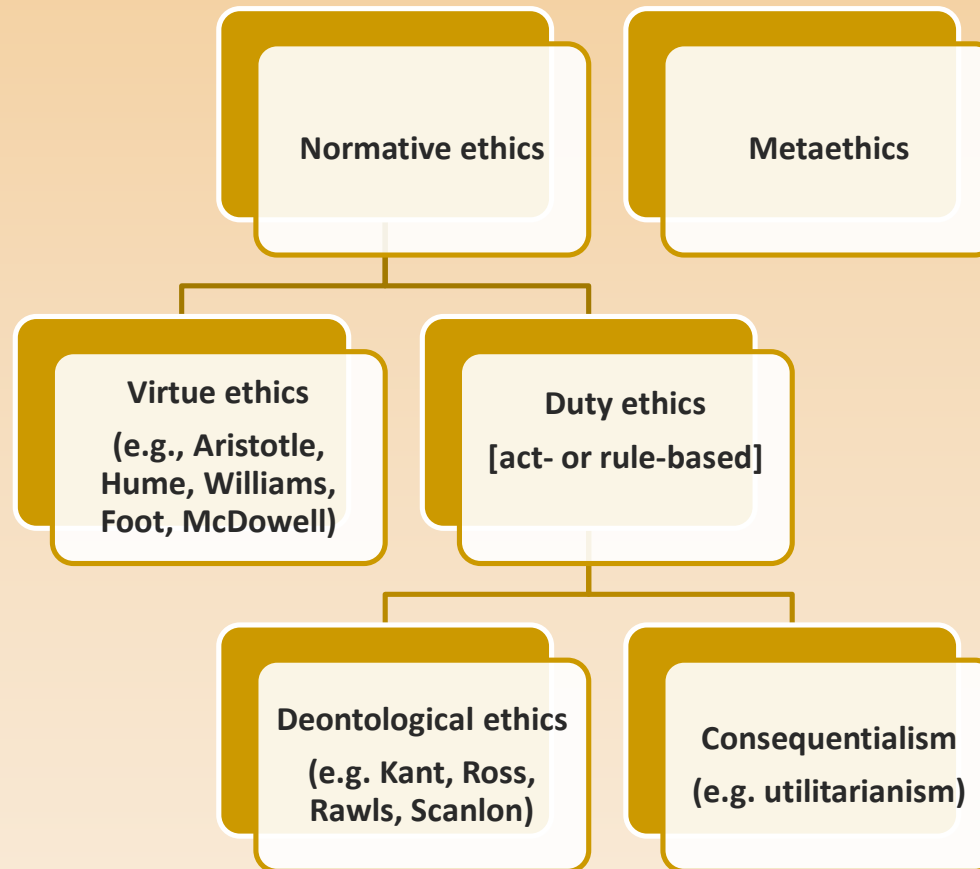
Foot on categoricity

- In setting up his alternative interpretation of Mackie, Joyce modifies Foot's early work on categorical vs. hypothetical imperatives.
 - Foot herself starts out treating oughts derived from etiquette, club rules, and the like as “non-hypothetical” insofar as they wouldn't be withdrawn [by those who accept the norms they presuppose] in application to agents who are indifferent to the norms.
 - But she shifts to accommodate a Kantian interpretation of them as hypothetical insofar as they need support from something further to have any reason-giving force (p. 315, second column).
- Her essential claim is that, however we classify them, their distinction from moral oughts can't be maintained. Kantian talk of “inescapability,” and the sense in which we “have to” act morally (= “the fugitive thought”), corresponds to nothing more than a *feeling* of compulsion.

Virtue ethics

- Foot's proposed alternative is a version of virtue ethics – one of the naturalistic approaches that Mackie claims is inadequate to capture the authority of morality.
 - The virtuous person is essentially an agent who's characteristically motivated by certain ends, e.g. justice, charity.
 - What Foot later "recants" (in a note tacked on to the article) is just her earlier assumption that [practical] rationality doesn't require moral ends, or concern for the good of others, which she attributes to her earlier acceptance of a Humean view of practical reasons, as dependent on agents' contingent ends.
- Joyce will accept a modified version of the Humean view of reasons, drawing from Williams's defense of "reasons-internalism, along with Mackie's Kantian picture of morality as requiring categoricity.

Basic normative ethical alternatives



Joyce on Mackie's argument

- Besides establishing problems with taking MI as T (the non-negotiable thesis), Joyce claims at the beginning of ch. 2 that Mackie's argument from queerness can't really rest on it, since it applies to the agents who believe moral judgments, not to the properties the judgments describe.
- [But can't the latter be taken as properties such that (even illusory) awareness of them motivates? That sounds "queer" enough.]
- Joyce also argues briefly that MI would be true in all possible worlds, whereas the argument from queerness pertains only to the actual world.
- [But Mackie's "Where in the world...?" can be taken as a rhetorical device – and it's unclear how Joyce's ensuing argument from reason-giving force could be limited to the actual world, since his "translation test" would also apply to inhabitants of other possible worlds.]

Joyce on inescapability

- Joyce thinks Mackie's argument from queerness really must rest on the "bindingness" or inescapability of morality: that it has jurisdiction even over agents who don't make moral judgments and have no desires that are furthered by behaving morally (e.g. Plato's invisible Gyges).
- He counts Foot's nonmoral institutional oughts as categorical but "weak." They're "non-evaporable," applying even to agents who have no interest in participating in the institution, e.g. the unwilling gladiator Celadus, but they fail to supply such an agent with ("real") all-things-considered reasons ("of his own").
- What distinguishes moral oughts as "strong" is supposed to be the fact that they give *everyone* reasons – though Foot's article dismissed this Kantian claim and instead held that moral oughts were hypothetical, i.e. applicable in virtue of agents' shared desires and interests.

Normative internalism

- Joyce replaces MI with the inescapability of moral *reasons*, contrasting Mackie's view with Kant's and Foot's (see chart, p. 38) – on the substantive and the conceptual claims of error theory, respectively.
- According to Mackie “ought” by definition *implies* “has a reason” (cf. MP = “Mackie’s Platitude,” p. 38 – and note that this is not an alternative to MI, i.e. a version of T, but just introduces something implied by T).
- [Note too that Joyce seems to think this entails having a *further reason* -- a reason *why* one ought -- that would “engage” the criminal, e.g. (p. 44).]
- But for an agent like Celadus, who stands outside an institution or system of norms, the reasons it offers to insiders, and that they can correctly ascribe to him as applications of the rules, don't amount to “his” reasons.

Institutional vs. “real” reasons

- Joyce distinguishes institutional from “real” reasons (= reasons an agent “can’t legitimately ignore”).
- He supports Mackie’s claim with the case of an agent who’s completely unaware of some arbitrary scoring system that people he doesn’t know are using to assess him (p. 41).
- On Joyce’s view the source of “real” reasons will turn out to be practical rationality.
- So “legitimately” will essentially turn out to mean “rationally” – though Joyce’s treatment of practical rationality will indicate that he takes it to include, not just prudential reasons, but any all-things-considered reasons, including overriding moral reasons (if any).

Joyce's categorical imperatives

weak

= institutional

give ("real") reasons only to insiders – who can apply them to [certain] outsiders

strong

= binding, inescapable

give ("real") reasons to everyone, regardless of desires or interests

Setting up Joyce's argument

- Joyce own argument for error theory, as sketched in basic form on p. 42, turns on the claim that it makes no sense to ascribe reasons to an agent that are independent of his desires or interests.
- Against Carnap's dismissal of "external" questions about the adequacy of a given framework as pseudo-questions, Joyce notes that the moral framework is nested in the wider framework of practical rationality.
- But practical rationality isn't itself an institution we can step out of. In raising questions about it, we'd be asking for further reasons, so the questions would be internal to it, after all (p. 47).
- Humean instrumentalism takes promotion of the agent's actual ends as the criterion of practical rationality, but Joyce's next chapter will defend a somewhat broader version of instrumentalism.

[A problematic point?]

- Here's a quotation from p. 62 that illustrates one worry I've had about Joyce's transitions back and forth between the questions of non-evaporability and universal reason-giving force of categorical oughts:

If you are told that somebody named "Jack" broke into a stranger's house, attacked the inhabitants, and all with the intention of taking their money for idle purposes, then from a moral point of view you have all the information needed to condemn Jack's action. There is no need to investigate further his desires (or ends); the judgment is not made under the assumption that his desires were a certain way.

- Didn't Joyce present our condemnation of the criminal (and the Nazi, and Gyges) as a matter of the nonevaporability (vs. inescapability) of a categorical ought? He then used Foot's argument to show that the same is true of "weak" categorical oughts such as club rules. Reason-giving force then came up as a further requirement of "strong" categorical, i.e. moral, oughts (and later will be linked to desires, on an instrumental account of reasons). But is that enough to make it a requirement for condemnation of outsiders by insiders?]

Williams on internal reasons

- Williams *understands* “internal” reasons (roughly) as reasons the agent has some motive for satisfying. He ultimately rejects the contrasting category, of “external” reasons.
- A simple “sub-Humean” model for the internal interpretation of reason-statements (p. 363) would make out the actions in question just as satisfying the agent’s actual desires.
- Instead, to allow for error about one’s reasons (cf. his gin/petrol example) and avoid restriction to desires specifically to do the act in question, Williams characterizes internal reasons in terms of four propositions (see pp. 364f.) that make them dependent on the agent’s “subjective motivational set” (S) while ruling out dependence on false belief.

Against external reasons

- A nonmoral example of the contrasting category of external reasons would be provided by reasons relative to *needs* – e.g. to take a certain medicine to preserve one's health – that fail to connect with a particular agent's desires. But we'd probably try to make sense of this by filling in desires besides those the agent acknowledges.
- For another example of a putative external reason, Williams cites Owen Wingrave's father's insistence that the family's military tradition provides a reason for Owen to join the army, even though Owen has no supporting motive for joining.
- Williams rejects this on the ground that a reason has to be capable of explaining motivation. It could do so only if the agent came to *believe* it was a reason, but then the disposition to act on it would have become an element of S, so the agent *would* have an internal reason, after all.

Application to moral reasons

- Williams's argument is meant to debunk Kantian attributions of moral reasons even to those with no supporting motivation – though he avoids comment on categorical imperatives or the relation of reasons to oughts.
- He allows that negative virtue-ethical terms (e.g. “selfish”) might be applied truthfully to agents like Owen who act in ways one thinks they should not.
- But he assumes that external reason statements are meant to convey a charge of *irrationality* for failure to act, and so they're false. The charge amounts to mere “bluff.”
- [Maybe the connection between reasons and rationality should be questioned?]

[Joyce on rationality]

- In ch. 3 Joyce sets out to argue against those who see practical rationality as yielding “real” moral reasons, independent of the agent’s ends. Instead, he defends a non-Humean form of instrumentalism (= reasons as prescribing means to the agent’s ends) for *normative* practical reasons.
- He first allows for alternative reasons in cases like Williams’s gin/petrol case by distinguishing Williams’s objective reasons (to do what actually furthers one’s ends, on Joyce’s instrumentalist reading) from subjective reasons (those the agent is justified in believing to be objective reasons).
- Rationality amounts to being guided by one’s *subjective* reasons, since irrationality entails arriving at beliefs or actions in a faulty manner, for which a [normal] person is responsible and deserves censure.

[Vs. Humean instrumentalism]

- For Joyce's purposes, "ends" can be construed vaguely to cover both desires and interests (e.g. smoker Wanda's interest in not smoking [contra Williams]), though Hume restricts it to the agent's present desires.
- Hume denies the possibility of practical rationality or irrationality and assumes that we always act on our [strongest] present desire.
- However, Joyce notes that intuitively it does seem that we can have and act on *irrational* desires. Rage and other emotional states can keep us from desiring what we think are the best means to our ends.
- Joyce's central argument in ch. 3 sketches an instrumental alternative to Hume, drawn from Michael Smith's work, that makes out our ends as themselves subject to rational appraisal.

[Vs. [neo-]Humean morality]

- Some contemporary Humeans try to defend a version of morality based on hypothetical imperatives by trying to identify ends all humans have, such as self-interest and/or sympathy.
 - But self-interest won't always support morality in particular cases.
 - Moreover, as Hume recognized, sympathy isn't something we always feel, or feel for *all* fellow humans, as opposed to our inner circle.
- Foot instead makes out morality based on hypothetical imperatives as escapable by those who don't share the relevant ends [though non-evaporable]. But Joyce thinks that doesn't fit ordinary moral discourse.
- Cf. Joyce's Nazi case, p. 62. [But does condemnation imply an attribution of reasons to someone whose desires don't support morality?]

Smith's instrumentalism

- Michael Smith argues from a non-Humean theory of reasons to a rationalist version of moral realism, but his argument rests on many questionable claims, e.g. MI* from ch. 1 and the claim that we “expect” a rational agent to act morally.
- However, Joyce agrees that we need a non-Humean theory of reasons to handle weakness of will – which Smith makes out as acting contrary to our *values* = those desires (not necessarily our strongest, as in Hume) that would survive rational deliberation meant to unify our initial stock of desires (see pp. 71f.).
- Joyce also follows Smith (among many others) in distinguishing *normative* (justifying) reasons from *motivating* (explanatory) reasons.
- He takes “*rational reasons*” as based on desires that would survive rational deliberation – vs. *irrational reasons*, capturing the remaining Humean reasons; cf. Molly’s case, pp. 72f. and p. 74.

[From rational to moral reasons?]

- Smith's own notion of rationality (like Brandt's, critiqued earlier) implausibly requires possession of all true beliefs.
- However, to test Smith's claim that subjectively rational reasons [= values] yield reasons with the content of ordinary morality, Joyce ultimately has us consider an idealized version S^+ of agent S who gives S rational advice (cf. Railton).
- *Whether or not* one attributes all relevant true beliefs to S^+ , it looks unlikely that there would be enough convergence in different agents' idealized desires, as needed to support morality – as Joyce will go on to argue in ch. 4.

The rationalist's dilemma

- A rationalist (= someone who views moral violations as irrational) faces a choice between
 1. letting [some] rational agents be alienated from [some of] their acknowledged normative reasons, and
 2. making such reasons relative to a particular agent's desires
- Smith manages to avoid 1 by basing morality on an ideally rational version of *the agent* (vs. Firth's Ideal Observer).
- But avoiding 2 then requires the assumption that full information and careful reflection would eliminate [morally relevant] differences in rational agents' desires.

Against Smith

- Joyce thinks it would be incoherent to hold that rational agents could *ever* be alienated from what they regard as normative practical reasons.
- [Is this really so obvious? Both Smith and Joyce seem to assume a Williams-style basis of reasons in an agent's own desires or other motives, but there are alternatives phrased in terms of promotion of value (cf. Raz on reason to read Proust).]
- Joyce undermines Smith's brief defense of rational convergence (see Joyce, pp. 88ff.) with alternative explanations of cases where agreement is reached, e.g. mutual social influence and/or the need for cooperation, rather than rational argument [alone].

A case from Harman

- Harman comes up initially as the source of a case that counts against Smith's insistence on rational convergence (even with a "Kantian leap"): a mobster (later named "Al"), employed by Murder, Inc., who stands to gain financially by murdering a non-family member, Bernard Ortcutt.
- Harman's point is that we wouldn't be providing the mobster with a reason by appealing to moral considerations, though he isn't in any way irrational. [Cf. Williams on Owen Wingrave.]
- Joyce takes this to illustrate his claim that instrumental reasons are agent-relative. But since moral reasons are conceptually non-relative, Smith must be wrong in claiming they're derivable from instrumental reasons. So moral rationalism is false – whereas other conceptions of morality [e.g., virtue ethics] lack the element of authority.

[Vs. Harman's moral relativism]

- Harman's own relativistic moral view represents another alternative to error theory insofar as it's designed to *save* ordinary moral discourse – by making out moral judgments as elliptical for the agent-relative judgments that *are* derivable from instrumental reasons.
- But Harman's analogy to the way we save pre-Einsteinian judgments of motion in light of relativity isn't apt, Joyce says, because we're not accommodating a new discovery. Moral relativism has long been recognized and denied by ordinary moral discourse.
- We'll be *abolishing* moral discourse, on the model of phlogiston or witch discourse, if we reject absolutism. (Cf. Joyce's Nazi case, pp. 98f.)

Using Williams's argument

- Williams takes all reasons as internal, meaning that they would motivate a fully informed version of the agent who deliberated correctly.
- Joyce limits his argument against external reasons to normative (as opposed to institutional) reasons, using Molly's case from ch. 4 to illustrate the various types of reason (among other things, distinguishing objective from Williams-style reasons, involving a choice made in light of full information; see pp. 107f.).
- The argument rests on both the Humean assumption that motivation requires a desire (not just a belief) and an understanding of reasons as potentially explaining action. (Cf. Joyce's argument against "besires," with mixed directions of fit.)
- Joyce goes on to defend Williams's argument against objections, on the basis of Williams's own conception of reasons.

Sketch of replies to objections

- Millgram's example of a reason – to see a therapist in order to become more sensitive – that Archie doesn't care about but would find satisfying to act on could explain action by Archie's becoming increasingly like Archie+, deliberating better on fuller information, while he still has a reason for improvement. But then the reason would be internal, after all.
- Hampton confuses the (false) claim that we always have an external reason to deliberate correctly with the claim that we always ought to act *in accordance with* correct deliberation (i.e., take S+'s advice). In any case, her external reason wouldn't yield reasons with moral content.
- Korsgaard's insistence on agent-neutral reasons confuses agent-relative with private reasons, misapplying Wittgenstein's "private language" argument. Valuing his own humanity needn't supply Murder, Inc.'s "Al" with a reason to value others', unless that gains support from his desires.

Joyce's qualified conclusion

- That completes Joyce's argument for moral error theory – which he now limits to establishing that it's *reasonable to believe* .
 - He notes that he doesn't take Williams's argument to rule out all external reasons, since institutional reasons may be external.
 - Instead, the argument cuts against *normative* external reasons (= irrational to ignore?): reasons that transcend all institutions, as in the case of practical and allegedly also moral reasons (see p. 134).
- The burden is now therefore on rationalists to show how their claim that morality possesses a distinctive kind of authority is anything more than bluff.

[Not the naturalistic fallacy]

- Moore's "naturalistic fallacy" is often misinterpreted by evolutionary theorists as a version of Hume's "is/ought" argument applied to "good," but instead it concerns the *definability* of the concept.
- The fallacy amounts to inferring identity from predication [e.g., deriving a definition of good from the claim that a certain sort of thing, e.g. pleasure, *is* good]. But naturalism (defined on p. 154) needn't be based on that confusion.
- Even if it's not based on a fallacy, though, naturalism doesn't seem to yield an adequate account of moral discourse. The strategy of revising the meaning of a concept to fit its extension would also vindicate discourse about witches and the like (cf. p. 156).
- Just as supernatural powers are essential features of witches, the whole point of moral discourse is its categorical force. [But cf. social coordination – and other possible defenses of the moral right to punish violators.]