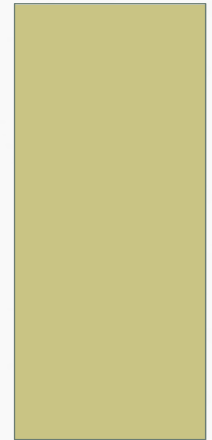# MORAL PSYCHOLOGY

PHIL 408P

STUDENT DATA (ON CARDS)

- <u>Contact info</u>: Please fill in your name, address, phone number, university ID, etc.

- Make sure the registrar has your correct email address (the one where you read your email regularly) so you'll get my "coursemail."

- <u>Background</u>: Please give names of courses, not just numbers; or indicate subject matter, if you forget the names.

- Note any courses that were taken elsewhere, or in other departments, and any taken concurrently with this course. If you have a particular reason for registering for this course, note that too.

WEB MATERIALS

- I keep my syllabus short by putting general policies and the like on the "courses" page of my website, whose address appears on the syllabus.

- The page contains a link to the syllabus, with further links to the schedule of assignments and other course materials such as supplementary readings.

- I don't use Canvas, but instead put relevant material on my website. I'll post the slides I give in class once a week, after our Thursday class – but it's dangerous to conclude that reading the slides is an adequate substitute for attending class.

RECOMMENDED BACKGROUND

- This course was supposed to have a Departmental prerequisite of PHIL 341 ("Ethical Theory") plus another prior upper-level course in philosophy – or consent of the instructor – but things got messed up when a graduate student taught his own version.

- PHIL 341 or the equivalent is still recommended, but the only prerequisite is two prior courses in philosophy. However, some of our courses don't really include any material on philosophical ethics (or even philosophy).

- If you don't have background in historical figures in the history of ethics, be aware that you'll have to do some further reading to understand our authors' references to such figures as Kant and Rawls.

- To test your readiness for the readings, you might sample some of the excerpts from Rawls that I've put on my website.

READINGS AND REQUIREMENTS

- Our main readings will be drawn from a book of essays and a collection of articles illustrating two very different current approaches to moral psychology: *philosophical* and *empirical.*

- The required articles discussing empirical work are all available on my website in their original form, so buy the recommended anthology only if you want to read shortened versions of some of them or to explore further work in the area, e.g. for our oral reports in the last third of the course.

- I've also recommended the textbook used by the graduate student who taught last year's version of the course, though it's heavily weighted toward the empirical approach, i.e. "cognitive science."

- Written requirements of the course include midterm and final exams: brief essays and/or identifications, chosen from a list that I'll distribute at least a week earlier.

INITIAL SUBJECT MATTER

- First we'll read three related essays by Nomy Arpaly who's roughly in the tradition of Aristotle and Hume, focusing on questions about the moral worth of persons, their motives, and their actions.

- However, she does grant that a full account of moral worth depends on an account of the rightness of acts, and in that respect she differs from virtue ethicists.

- She also departs from Aristotelian virtue ethics by not stressing general character traits, which are disputed in some recent empirical work (e.g. by John Doris, whom she cites).

- Instead, she stresses *motivation* – and argues that emotional motivation to act *against* one's "best (or better) judgment" can be both rational and moral – sometimes more so than the likely results of reasoning or deliberation (as on the Aristotelian model assumed by most contemporary action theorists).

FIRST ASSIGNMENT

- For next time, read Arpaly's first chapter, which sketches various types of cases that she thinks standard philosophical approaches to action (e.g. by Donald Davidson) don't handle adequately. Then at least get started on ch. 2, about the rationality of action against one's better judgment.

- You might choose a case or two that you're inclined to question and/or find interesting. If she's arguing against an approach you're not familiar with, and you don't get a clear enough idea from her explanation, bring it to my attention in our next class, and I'll try to explain it further. But look up any unfamiliar terms online before then.

- Also, locate the recommended reading by Davidson on my website, for an idea of the approach she's questioning – and to be sure you know how to access that material. The username and password you need to use is my last name, "Greenspan" (with a capital "G"), as noted on the syllabus.

LATER COURSE TOPICS

- In the second third of the course, we'll take a critical look at recent interdisciplinary work by psychologists and empirically-minded philosophers – also concerned with emotions, but in a very different way: as the source of moral *judgments*.

- Some of the philosophers, e.g. Singer, argue that the "intuitions" many philosophers such as Rawls rely on in constructing or criticizing moral theories are unreliable, since they're typically based on evolved emotional responses rather than reasoning.

- However, we'll also take a look at Rawls's own initial description of his method – and the work of a psychologist who views the emotional basis of moral judgment positively (Haidt). I'll point out ways in which Singer and others may be misinterpreting these authors.

INTERACTIVE ORAL REPORTS

- In the third part of the course, we'll form groups to lead discussion of special topics.

- I refer to these student-led discussions as "interactive oral reports," because I play an active role, raising questions, making occasional corrections, and sometimes just slowing you down so we can digest the material. I invite other students to do so as well, so you shouldn't plan these as solo presentations.

- You might want to discuss some of the source material in the recommended anthology or text. If you can scan any further readings, I'll post them on my website, so that everyone can have a copy.

MORAL AGENCY

A PHILOSOPHICAL APPROACH

STANDARD MODELS OF ACTION

- Arpaly discusses cases are supposed to cast doubt on philosophers' usual model of *practical rationality* (= rationality concerning action).

- It's a model of *deliberation*, i.e. conscious reasoning resulting in a judgment about what one ought to do. The latter is referred to as one's "better (or best) judgment," to the extent that it purports to be based on consideration of all available evidence.

- The model stems from Aristotle's "practical syllogism," which was put forth in contemporary form by Donald Davidson in his influential article on weakness of will, construed as acting against one's better judgment – or in the Greek term philosophers use, *akrasia* (vs. *enkrateia*).

- I've listed Davidson's article as "recommended" and put a copy on my website, in case you'd like to see his account in detail. Or just take a look at p. 33, for Aquinas's explanation of akrasia in terms of competing practical syllogisms.

ARPALY'S MAIN DEPARTURE

- In an influential article, included in her book as ch. 2, Arpaly argues for what she calls "*inverse* akrasia" – a concept that will come up repeatedly in discussion of the cases in ch. 1.

- Inverse akrasia "inverts" the usual assumption that it's irrational (and in moral cases, wrong) for agents to act against their better judgment.

- Arpaly's claim is that agents' responses to emotional cues are often better guides to action than their "better judgment," given that their deliberation may have omitted important considerations. In quite familiar cases, then, it's rational to be *akratic*, or weak-willed.

A MUCH-DISCUSSED CASE

- Arpaly has various examples, many quite original, as you'll see – including an unusual number of examples involving antisemitism (she's originally Israeli) – but one case raising issues of moral worth or praiseworthiness that she shares with other philosophers is that of Huckleberry Finn in Mark Twain's novel.

- Huckleberry Finn's "conscience" – what he'd learned from his elders in an environment where slavery was the moral norm – tells him he ought to turn in Jim, a runaway slave.

- But he can't bring himself to turn Jim in, having gotten to know him as a person. He acts against his better judgment, and he's ashamed of being weak-willed – though he thereby does what in fact is the right thing.

- On Arpaly's account in ch. 3 he displays more moral worth than he would if his conscience were his guide. Her arguments in ch. 2 also suggest that he's acting rationally.

TYPES OF CASES IN CH. 1

- <u>Compromised autonomy</u>: passivity with respect to one's own morally significant actions; cf. Le Carré's Oliver Single. *Six more*:

- <u>Misguided conscience</u>: inverse akrasia; cf. Peter the compassionate Nietzschean, Huckleberry Finn – vs. the child-beater, the Nazi, etc.

- <u>Self-deception</u>: epistemic (= theoretical) irrationality motivated by desires ("hot") vs. Davidsonian intentional act; cf. anorexic's self-image, Voltaire's beliefs about Jews, 19th-c. doctors' beliefs about masturbation, etc.

- <u>"External" desires</u>: desires from which the agent feels estranged ("not me" – or "true colors"?); cf. "ego-dystonic" homosexual, "unwilling addict" (Frankfurt)

- <u>Unconscious motivation</u>: ignorance of one's real motives vs. philosophers' stress on first-person perspective; cf. anorexic, "fame-chaser"

- <u>Nondeliberative inference</u>: reaching practical conclusions without conscious reasoning (vs. Smith, Korsgaard); cf. reaching for nuts, picking up on facial expressions, inverse akrasia

- <u>Nondeliberative rationality</u>: nondeliberative conclusion more rational than competing deliberative conclusion; cf. Tamara debating whether to marry Todd (on the assumption that her emotions are more attuned to her reasons).

NOTIONS OF RATIONALITY

- **Arpaly notes that Tamara's r**ationality in deciding (or deciding not) to marry Todd depends on whether her decision is responsive to the relevant reasons, not on the state of mind in which she arrived at it (e.g., emotional upset; see p. 27).

- Though Arpaly doesn't give a definition of "rational," which is a disputed term among philosophers, what she means by the term emerges gradually, as she contrasts it with other possible interpretations.

- Four common interpretations emerge in her discussion in ch. 2, the fourth of which is hers:
  1. Reached in a calm, "deliberate" state of mind
  2. Producing the best outcome
  3. Mentally coherent/consistent
  4. Responsive to (good) reasons

A THIRD-PERSON ACCOUNT

- One of the main features of standard models that Arpaly calls into question in ch. 1 is the emphasis on the first-person perspective, the perspective of a deliberator.

- In ch. 2 she begins by distinguishing an account of rationality from a deliberator's manual giving advice about how to act rationally.

- Though you can advise someone to act against what *in fact* is her best judgment, e.g. tell Tamara not to marry Todd, "Act against your best judgment" would be an absurd piece of advice, since convincing Tamara to act on it would make *that* her "best judgment."

- However, what Arpaly is aiming to provide is an *account* of rationality, which works from a third-person perspective.

STARTING POINTS

- Arpaly defines "best judgment" as "the judgment that *one reaches*, having taken into account all the reasons *one judges* to be relevant, as to what would be best for one to do in a given situation" (p. 35).

- She notes that her argument for the rationality of inverse akrasia assumes that reasons are based on the agent's desires and beliefs, so it doesn't depend on the existence of "external" reasons [in Bernard Williams's term].

- The standard view of akrasia as irrational is usually just taken for granted without argument, following Davidson in his account of weakness of will. But recently Michael Smith has defended a version of the standard view against objections, so Arpaly begins by criticizing his argument.

- Smith's claim is that, if you believe you should φ -- which he interprets to mean that you believe you'd want to φ if you were rational -- then you rationally should *desire* to φ.

THE APPEAL TO COHERENCY

- Dreier and Sayre McCord object that Smith's claim would mean that whatever you think it's rational to do, even if that belief is irrational, would be rational to desire to do – and assuming that the desire isn't opposed by some stronger desire, rational to do. So believing something rational would make it so!

- Smith replies by granting that the desire based on an irrational belief wouldn't be *fully* rational, but maintaining that it would be *more* rational than not having the desire corresponding to one's belief.

- Smith says that the coherency and consistency of an agent's mental states is what rationality is about, rather than finding the optimal outcome, and a conflict between belief and desire would undermine it.

- But Arpaly argues that taking account of one's *overall set* of beliefs and desires, not just a single belief-desire pair, would show that it's sometimes more rational to (desire to) act against one's best judgment (= belief).

THE CASE OF THE FAILED HERMIT

- Arpaly's main example is Sam, a student who's anxious about not getting enough work done and therefore deliberates to the conclusion that he ought to become a hermit, in the sense of cutting off all social contact.

  - But Sam never manages to generate the requisite motivation to become a hermit and hence acts against his best judgment.

  - In fact, though, becoming a hermit would have made Sam too depressed to get much work done, so he's acting rationally.

- Arpaly's claim is that, in light of other beliefs and desires that Sam has but fails to consider when deliberating (e.g., his memories of becoming depressed without social contact and his desire to be happy, p. 42), a desire to become a hermit would *lessen* the coherence among his mental states.

- It's his belief that he ought to become a hermit that introduces incoherency into his mental states, and adding a corresponding desire would just compound it.

SMITH'S QUALIFICATION

- Smith responds by limiting his point to deliberation based on *evidential* considerations – whereas Sam's deliberation is said to be based on anxiety.

- But Arpaly asks us to consider the contrasting case of Paul, whose much more reasonable judgment that he ought to devote three hours a day to his studies is also prompted by anxiety, bypassing deliberation entirely, though his decision to act accordingly does seem rational.

- Here again we can't equate rational decision with a decision reached in a certain (calm, "deliberate") state of mind.

- The problem with Sam's deliberation is that it doesn't take account of all relevant evidence, particularly evidence about his own emotion tendencies, not that it was prompted by emotion *rather than* evidence.

REASONS-RESPONSIVENESS

- Some object that action against one's better judgment would have to be produced by an irrational process.

- A paradigmatic case is that of Nicole (p. 47), who doesn't recognize that she lacks the talent to become a professional singer but instead fails to act on that aim because of irrational inhibitions derived from a childhood that undermined assertiveness in achieving her own aims.

- But this fits only some versions of Sam's case. We might suppose instead that Sam continues his social engagement for the very same reasons that he overlooked in his deliberation (p. 48), rather than out of laziness or the like (as he thinks).

- See pp. 49f. for the cases of Emily vs. Alice, who leave their grad program in chemistry for good vs. bad reasons. The question is whether their act is triggered by the good reasons both of them share, for thinking they wouldn't be happy in the field. Alice's act seems instead to be triggered by low self-esteem.

SUMMARY: CONTRASTING CASES

- <u>Sam</u>: deliberation defective; anxious, but acting in response to relevant reasons [for not becoming a hermit], akratic act rational

- <u>Paul</u>: also anxious, but arrives at a reasonable judgment on that basis [that he ought to spend three hours a day studying], so acting on it is rational. Shows that Sam's anxiety doesn't rule out rationality.

- <u>Nicole</u>: deliberation defective, but akratic act motivated by irrational inhibition vs. relevant reasons, so irrational. Treated by other moral psychologists as the standard case of akrasia.

- <u>Emily</u>: deliberation defective; acts out of unease that responds to relevant reasons [for leaving chemistry program], so akratic act is rational.

- <u>Alice</u>: deliberation also defective, but acts out of low self-esteem, rather than relevant reasons, so akratic act is irrational.

NONDELIBERATIVE CASES (1)

- Arpaly goes on offer three sorts of cases in which one can see that deliberation isn't necessary for rationality.

- The first is *fast action*, e.g. an accomplished tennis player's decisions during a game, or those of a witty participant in a fast-paced conversation, contexts in which one can't take time to deliberate.

- We often praise such agents for rational choices ("brilliant") – and occasionally criticize their choices as irrational ("crazy," "foolish").

- An appeal to an earlier deliberate decision to trust one's instincts doesn't fit all cases (e.g. James, who leaves a room without knowing why but is responding to perception of hostility).

- It also would not make sense of our judgments of irrationality in such cases, where the agent's instincts turn out to be triggered by *bad* reasons (e.g., James leaving because he's reminded of his bad-tempered uncle; or certain choices made by the tennis player or witty conversationalist).

NONDELIBERATIVE CASES (2)

- Certain cases of *belief formation* are commonly judged rational without a basis in deliberation.

- The simplest example is of beliefs formed on the basis of perception, e.g. Arpaly's judgment that there's a black cat in front of her.

- For more complex examples consider "dawning," or nondeliberate formation of belief on the basis of gradual exposure to new evidence, e.g. Candide's realization in Voltaire's novel that this is *not* "the best of all possible worlds."

  - Or cf. an alternate version of Emily's case, where she realizes before leaving her grad program that chemistry is not for her.

  - Not all dawnings are rational, though: cf. cases where a desire prompts a change of mind or a failure to change one's mind without reference to evidence, e.g. accepting Jesus just to avoid feeling alone ("wishful thinking").

NONDELIBERATIVE CASES (3)

- Another case comes from considering the act of *deliberation itself:* the decision *to* deliberate can't always be based on deliberation, but that doesn't make it irrational.

- Otherwise, we might have a regress: deliberation about whether to deliberate would be rational only if backed up by yet more deliberation – and so on *ad infinitum*.

- Arpaly gives the example of deciding what books to assign a class. Her deliberation starts from a limited list of possibilities that's rationally generated without deliberation from memories of her previous students' abilities.

- She notes that neurophysiologist Antonio Damasio has presented cases where the ability to reason practically is undermined by brain lesions resulting in emotional deficits (e.g. failure to feel uneasy about options that didn't work out in the past), even where the ability to deliberate is unimpaired. [Arpaly may be misinterpreting Damasio here; his cases show that emotions are *essential to* correct deliberation.]

CONCLUSIONS AND LIMITATIONS

- Arpaly grants that standard theories of rationality will be compatible with her argument, as long as they satisfy certain conditions (see pp. 60f.): e.g. not privileging "best judgment" or identifying reasons with those the agent is aware of *as* reasons.

- Her view does allow that deliberation has a point, but it's mainly useful in dealing with new information and unfamiliar data (p. 64). [Cf. also acts you think you'd otherwise perform badly. Other cases?]

- She notes that her main aim is to capture the moral worth of actions and agents – the subject of her next chapter.

- [Question : What would she say about Huck Finn, if the question is whether he's rational?]

[SOME GENERAL REFLECTIONS ON ARPALY'S RATIONALITY ARGUMENT

- **Essentially, on Arpaly's account, a rational act amounts to one that's justified for a particular agent by reasons to which he has access.**

  - **This is to say that the agent has access to facts relevant to choosing the best act (later: "de re" reasons), though he may not recognize them as such (in which case, they're not "de dicto" reasons).**

  - **Note that – as with justification and truth of belief – a rational act may not always be the right or best act for an agent to perform (in terms of rational appropriateness, not just morally).**

- **Note too that the main cases Arpaly's argument turns on (e.g., Tamara, Sam, Emily) seem to concern issues *involving* emotions, where "cool" deliberation might be particularly likely to miss relevant facts. Should her conclusion on rationality be limited to such cases?]**

MORAL WORTH

- As Arpaly sets up the problem of moral worth, using the example of Madame Bovary's charitable phase, it concerns how agents who perform the same acts can differ in praiseworthiness.

- She understands moral worth as praiseworthiness or blameworthiness, i.e. whether someone('s action) *deserves* praise or blame – vs. whether or not we actually ought to praise or blame her (action) under the circumstances.

- Moral worth depends on but is distinct from what Arpaly calls "moral desirability," which amounts to whether the act was right [but allowing for degrees (as moral worth does)].

WORTH VS. DESIRABILITY

- A famous case where moral desirability and moral worth come apart is Kant's prudent grocer, who prices goods fairly just in order to profit in business. The fact that he thereby does the right thing seems accidental, since the the reasons that motivate him have nothing to do with those that make his act morally desirable.

- So his act doesn't stem from moral reasons-responsiveness. In other words, his reasons for action don't correspond to the act's right-making features [= *de re* moral reasons].

- Arpaly presents a principle summing up her position so far (p. 72):
  <u>Praiseworthiness as Responsiveness to Moral Reasons (PRMR)</u>: For an agent to be morally praiseworthy for doing the right thing is for her to have done the right thing for the relevant moral reasons—that is, the reasons for which she acts are identical to the reasons for which the action is right.

DE RE VS. DE DICTO RESPONSIVENESS

- **Kant and Aristotle hold that moral worth depends on acting from a desire to do what's right: acting out of duty (for Kant) or for the sake of the noble, i.e. virtue (Aristotle).**

- **This amounts to acting from reasons *believed or known* to be moral [= *de dicto* moral reasons] .**

  - **But we can see that concern to act morally isn't sufficient for moral worth from Arpaly's example of "the Extremist" [in Israeli politics], who misidentifies the moral prohibition of killing someone as based [solely] on the fact that (s)he's a fellow Jew, rather than simply a person; cf. Ron's reason against killing Tamara.**

  - **That concern to act morally isn't necessary for moral worth is shown by cases of *inverse akrasia* that involve doing the right thing in response to moral reasons that one doesn't recognize as such – as in the case of Huckleberry Finn and other fundamentally good people who happen to incompetent abstract thinkers but whose acts are responses to right-making features rather than just squeamishness or the like.**

[LIMITATIONS

- Aren't there limits on Finn's moral reasons-responsiveness, as well as his judgment, unless he also judges *others'* acts in accordance with those reasons? How would he feel toward someone else's act of returning a runaway slave?

- Arpaly briefly grants that Huck's moral worth is limited, though she doesn't spell out why she thinks so. Is he really responding to *the* relevant moral reasons just by responding to Jim's status as a person? Mustn't he also believe that persons shouldn't be treated as property – and that other slaves are also persons?

- Arpaly goes on to consider blameworthiness and then *degrees* of moral worth, so we should ask whether her developed account would handle this issue adequately. Is Huck praiseworthy for some acts but blameworthy for others (i.e. acts of omission), or is his limitation a matter of degree – or both? Or something else?]

BLAMEWORTHINESS

- On a parallel account of blameworthiness for wrong action, it depends on either

  - responsiveness to sinister (essentially anti-moral) reasons (= ill will), or

  - insufficient responsiveness to moral reasons (= deficiency of good will)

- Arpaly's example here is Jeanne's rudeness to Joseph, distinguishing cases where Jeanne accidentally does wrong (doesn't realize that her remark would hurt his feelings), and hence isn't blameworthy, from cases corresponding to the categories above, where either

  - her motive was to hurt his feelings (*sadistic*) or

  - her motive was to vent her own tensions (*egoistic*), without adequate concern for how it would make him feel.

DEGREES OF MORAL CONCERN

- Other things being equal, sadistic action is worse than egoistic – contra **Kant**, who lumps them together as action from "inclination" (vs. duty).

- Arpaly notes that moral motivation needn't be proportional to moral perception (vs. "blindness"). The possibility of not caring much about morality is illustrated by Mary, the "half-moralist" (p. 82), who decides on nonmoral grounds not to fulfill her promise to mail a book to Russia today – in the snow, when she's facing a deadline.

- On p. 84 Arpaly offers an expanded version of PRMR, according to which one's degree of moral worth depends on the depth (= strength) of one's moral concern:

  Praiseworthiness as Responsiveness to Moral Reasons (revised version): For an agent to be morally praiseworthy for doing the right thing is for her to have done the right thing for the relevant moral reasons—that is, for the reasons for which the action is right (the right reasons clause); and an agent is more praiseworthy, other things being equal, the deeper the moral concern that has led to her action (the concern clause).

FEATURES OF MORAL CONCERN

- Arpaly characterizes moral concern as one's non-instrumental desire that (de re) morality be followed – noting that it's not the same as intensity of [occurrent] feeling, reflective endorsement [= approval of one's desire, as in Korsgaard], or commitment (reflective endorsement with some emotional backing).

- She describes three features associated with depth of concern, the first of which will be discussed in detail:

  - <u>Motivational</u>: Deeper concern has a "diehard" quality: it would take more to keep you from acting on your concern. [Might the second and third categories help to explain the limitations on Huck Finn's praiseworthiness?]

  - <u>[Affective]</u>: Deeper concern is reflected in stronger [dispositional] emotions, e.g. tendency to feel guilt at the thought of doing wrong, or anger at someone else for doing wrong.

  - <u>Cognitive</u>: Deeper concern involves more of a tendency to notice morally salient features of the situation.

COMPARING PHILANTHROPISTS

- To illustrate the "diehard" motivational quality of deeper moral concern, Arpaly contrasts three types of philanthropists:

  - <u>Sorrowing</u>: acts benevolently despite depression

  - <u>Fair-weather</u>: acts benevolently most of the time, though not when depressed

  - <u>Capricious</u>: acts benevolently (gives to charity) on a whim, just because it happens to be easy to manage

- All three philanthropists may be acting from moral concern, but the depth of their concern varies.

CONTRA KANT

- Arpaly's fair-weather philanthropist needn't be benevolent on ordinary occasions just because she has a cheerful temperament or tends to get pleasure from helping others, as Kant suggests.

- Even her capricious philanthropist may be acting out of moral concern, albeit only skin-deep.

- By contrast to Arpaly's sorrowing philanthropist, the only philanthropist Kant takes to display any real moral worth is

  - <u>Cold-hearted</u>: acts benevolently despite *lack* of feeling for others, just because it's his duty.

- Arpaly grants that Kant's philanthropist is underdescribed, and on some variants he might care deeply without feeling much. On others, though, his motivation would be compromised by conflicting motives or indifference.

"HALF-HEARTEDNESS" [?]

- Assuming that the moral concern of Kant's cold-hearted philanthropist is barely enough to motivate him, he's *half-hearted.*

- He might also be half-hearted in Frankfurt's sense, if his moral concern is countered by motives that essentially conflict with morality, e.g. contempt for others (p. 90).

- [In what sense does contempt "essentially" conflict with moral concern? What about a jaundiced view of humanity – not thinking highly of human nature, as was true of Kant himself?]

- Arpaly gives the example of someone committed to Puritan Christianity but subject to lust. ["Half-hearted" seems the wrong word here, since it implies weakened motivation, rather than intense ambivalence. Can't both motives be "deep," even if in conflict?]

IMPLICATIONS

- Inadequate concern for moral reasons also differs in degrees: Failure to do the right thing is less blameworthy if it doesn't depend on *much* indifference, i.e. if doing the right act would require an especially high degree of concern.

- Arpaly gives the example of not hiding Jews during the Holocaust: given the risks, it would take a high degree of moral concern to do and thus may not be very blameworthy to fail to do – even if it's not supererogatory but morally required.

- She also notes that inverse akrasia doesn't make one praiseworthy in cases of misguided conscience unless it reflects deep moral concern.

- She contrasts Huck Finn with Goebbels, who occasionally had attacks of compassion but was able to overcome them fairly easily. Supposing that on one occasion he went against his conscience to save a Jewish acquaintance, he might have acted from the same motives as Huck, but his concern was capricious, whereas Huck's (on Arpaly's reading) was deep.

THE RELEVANCE OF CHARACTER

- The moral worth of an action depends on the character of the agent, but on Arpaly's account character amounts to depth of moral concern, not the stability of a disposition to act morally – though the latter may be evidence of "diehard" motivation.

- She takes this to fit Aristotle's account of courage (bravery), which he understands as involving a willingness to sacrifice oneself for morally relevant reasons (overcoming fear of death with fear of disgrace).

- But the falsity of Aristotle's and traditional virtue-ethicists' assumptions about stable and unified character traits as essential for virtue has been shown by recent psychological research, as philosophers Doris and Harman argue.

- There's intuitive evidence too -- e.g. of a lack of connection between specific moral concerns that fall under the same general trait, such as courage in battle vs. courage in voicing unpopular opinions.

MISGUIDED CONSCIENCE

- Arpaly now turns to a more detailed discussion of the consequences of her account for cases of misguided conscience.

- Is there any merit in steadfast devotion to a wrong cause? Besides Goebbels, consider less sinister cases like the chastity promoter who thereby neglects worthier causes.

- An agent who's loyal to bad or foolish principles may exhibit self-control, but in itself that trait is morally neutral – amounting to a skill that can be used for bad ends as well as good.

- What we praise in other cases under the rubric of self-control is really the greater concern of the agent for morally relevant reasons over other considerations.

IGNORANCE VS. PREJUDICE

- Nor does Arpaly take misguided conscience involving false factual beliefs to absolve an agent from blame – e.g. if Hitler sincerely believed that Jews were conspiring to take over the world, etc.

- She deals with false belief by distinguishing "honest mistake" from "motivated irrationality," where the belief actually stems from sinister reasons or inadequate regard for moral reasons.

  - Her main example of honest mistake is Solomon, whose remote village environment keeps him from knowing about women's abilities as abstract thinkers.

  - Suppose Solomon gets a scholarship to college and is exposed to brilliant female abstract thinkers. If he still retains his belief that they lack that ability, he's prejudiced rather than ignorant.

ROLES OF FALSE BELIEF

- Arpaly does note at the end of ch. 3 that ignorance can qualify moral worth: e.g., Huck Finn's inability to tell his virtues from his vices may make him likely to try to make himself a morally worse person. [Is this an adequate answer to the problem I raised about limiting Huck's moral concern to his own action?]

- But sexism, racism, and other prejudices are blameworthy insofar as the beliefs they seem to depend on are really based on sinister motives (e.g., a desire to have someone to despise or hate or to feel superior to) or inadequate concern for morally relevant factors (e.g. expressing anti-semitism to try to fit in with friends).

- She distinguishes three versions of Hitler (pp. 100f.) and of the childbeater who acts from morally acceptable motives (p. 108). In her the third category, the false belief plays a real causal role, but it's motivated by a desire that's either sinister (Hitler) or outweighed by moral reasons the agent doesn't respond to adequately (the man who beats his child to sustain his love for abusive parents).

SORTING OUT CASES (1)

Arpaly's cases toward the end of ch. 3 are meant to illustrate various different claims, so it might be well to summarize some of them and their different points.

- Goebbels: akrasia results in morally desirable act, but not morally praiseworthy (vs. Huck Finn), since moral concern is shallow; capricious

- the chastity promoter: misguided conscience involving inadequate attention to more serious moral reasons

- Solomon: distinction between ignorance and prejudice (*not* misguided conscience)

SORTING OUT CASES (2)

- Arpaly also gives various cases of misguided conscience based on "motivated irrationality" rather than "honest mistake":

  - "sincere" Hitler

  - the anti-semite motived by desire to fit in with friends

  - the character-building childbeater motivated by desire to love parents

  - the snobbish colleague ("megalomaniac") who's really insecure

- These agents are still blameworthy, since their false beliefs are motivated either by sinister reasons (e.g., a desire to hate someone) or by inadequate attention to moral reasons.

[A GENERAL PROBLEM?

- Arpaly doesn't seem to say anything about how much weight one should place on different reasons, e.g. moral vs. self-interested reasons, when they conflict. But presumably moral reasons are stronger in at least some clear cases.

- If rationality amounts to responsiveness to reasons, however, that includes moral reasons. So if rationality involves responsiveness in proportion to the weight of one's reasons, it looks as though it *entails* moral praiseworthiness on Arpaly's account – contrary to the intuitive view that it's possible to be a rational evildoer.

- Many Kantians and at least some virtue ethicists (e.g., Philippa Foot) would welcome that view of practical rationality, but it's unclear whether Arpaly intends it. Is it plausible?]

MORAL JUDGMENT

EMPIRICAL WORK

QUESTIONING INTUITIONS

- Recently a number of philosophers have argued that recent empirical work undermines reliance on intuition as the basis for moral (along with other) judgments and theories.

- Moral intuitions, as the term is used by contemporary philosophers, following Rawls, means pre-theoretical opinions about right and wrong, etc. Arpaly relied on them in making her argument from particular cases to a view of moral worth.

- In a way, Singer reverses Arpaly's emphases, arguing that we should mistrust our intuitions in making moral decisions or constructing moral theories and instead rely on reasoning, even where its results seem counterintuitive.

SINGER'S NORMATIVE AIMS

- The more common appeal to moral intuitions that Singer wants to question is in arguments against utilitarianism.

  - Utilitarianism determines the rightness of acts just by how good their consequences are.

  - However, it seems to allow us to sacrifice one person to save more, as in "trolley cases."

- The contrast is to "deontological" theories, taking rules or principles of right/wrong as primary (e.g., Kant and Rawls).

- Rawls relies on intuitions as the basis for the method of ethical reasoning aimed at "reflective equilibrium." As described in his *Theory of Justice* this involves trying to bring our intuitions about general moral principles into balance with our intuitions about the rightness or wrongness of particular types of act by modifying one or the other.

SOURCES OF EVIDENCE

- Singer's argument against Rawls's method includes an overview of evidence undermining intuitions from three main sources. To sum them up briefly:

   1. <u>Evolutionary theory</u> (linking moral emotions to our evolutionary environment rather than current circumstances)

   2. <u>Psychology</u> (Haidt's experiments showing that ordinary moral judgment is formed intuitively and later "rationalized")

   3. <u>Neuroscience</u> (Greene's brain-imaging studies linking deontological judgments to emotion, whereas utilitarian judgments involve reasoning)

- We'll sample 2 and 3 in our further readings, along with work in the "heuristics and biases" literature suggesting that morally irrelevant factors (e.g., order of presentation) influence our intuitions.

EVOLUTIONARY BASES

- As an introduction to his central argument Singer describes how evolutionary theory confirms and extends Hume's description of our primary moral concern for people closer to us ("kin altruism") or those from whom we can expect return favors ("reciprocal altruism").

- Our sense of fairness/obligation/justice emerges from the need to distinguish reciprocators from nonreciprocators ("cheaters") and protect ourselves from the latter.

- Singer's claim is that, while evolutionary theory can't directly yield moral or other normative conclusions (since it explains rather than justifying behavior), it undermines some ways of doing ethics that themselves have normative conclusions, namely those based on intuitions.

SINGER'S CENTRAL ARGUMENT

Singer argues that evolved moral responses should be discounted, since they're keyed to the evolutionary environment and hence needn't fit contemporary situations (e.g., in trolley cases, killing by turning a switch vs. pushing over a bridge).

- He writes (p. 348): "What is the moral salience of the fact that I have killed someone in a way that was possible a million years ago, rather than in a way that became possible only two hundred years ago? I would answer: none."

- [But doesn't this answer cut both ways? It may be that our evolutionary "alarms" don't go off in contemporary situations where we *should* be goaded into paying more attention. Since Singer's own version of utilitarianism implies extending moral concern beyond our evolved feelings of sympathy to remote individuals, why not also extend our aversion to doing what Greene calls "up close and personal" harm (e.g., to nuclear bombing; cf. also Thomson's later view on trolley cases).]

RATIONAL INTUITIONS?

- **Ultimately, Singer distinguishes between emotion-based intuitions, as products of evolution, from "rational intuitions," which are needed to support utilitarianism.**

  - Greene's results on reaction time in trolley cases suggest that some people can overcome an initial intuitive aversion to "up close and personal" harm.

  - Singer grants that there's an intuition involved in the utilitarian judgment that the death of one person is "a lesser tragedy" than the death of five (p. 350), or even just that being killed is a bad thing. But he says this isn't a product of evolution [?].

- **Singer's general conclusion is that we need to separate our rational from our emotional intuitions [or those valid today from those keyed to an earlier environment? But would that justify discounting *all* emotion-based intuitions, in the way he claims is warranted? Does his argument just rest on assumption of emotion/reason dichotomy?].**

INTUITIONS IN PSYCHOLOGY

- One of the sources Singer and other philosophers cite in their argument against [emotion-based] intuitions is social psychologist Jonathan Haidt.

- Haidt is really arguing against the rationalist tradition in psychology. He cites a study indicating that only philosophers and others with "a high need for cognition" reach moral judgments by reasoning.

- The usual source is moral intuition, which he defines (p. 818) as

  ...the sudden appearance in consciousness of a moral judgment, including an affective valence (good-bad, like-dislike), without any conscious awareness of having gone through steps of searching, weighing evidence, or inferring a conclusion.

- He understands reasoning, by contrast, as having at least two conscious steps, and he assigns it only a "defensive" role in typical individual moral judgment.

REASONING EX POST FACTO

- Such "private" moral reasoning as people normally engage in, according to Haidt, occurs only *after* they've reached a judgment, and it exhibits "confirmation bias."

- Rather than weighing pros and cons, people ordinarily "confabulate" reasons for judgments they've already made ("rationalize" those judgments, in the psychologists' sense = make up spurious reasons for them).

- Their aim is "defensive" in the sense that they're trying to defend their image of themselves as mentally coherent – in the manner of a lawyer defending a case, rather than a scientist seeking truth.

- Reasoning is likely to have causal influence only on *other people's* intuitions and judgments, via social interchange – though its effectiveness tends to be limited to persuasion of others who aren't yet committed to opposing judgments.

"SOCIAL INTUITIONISM"
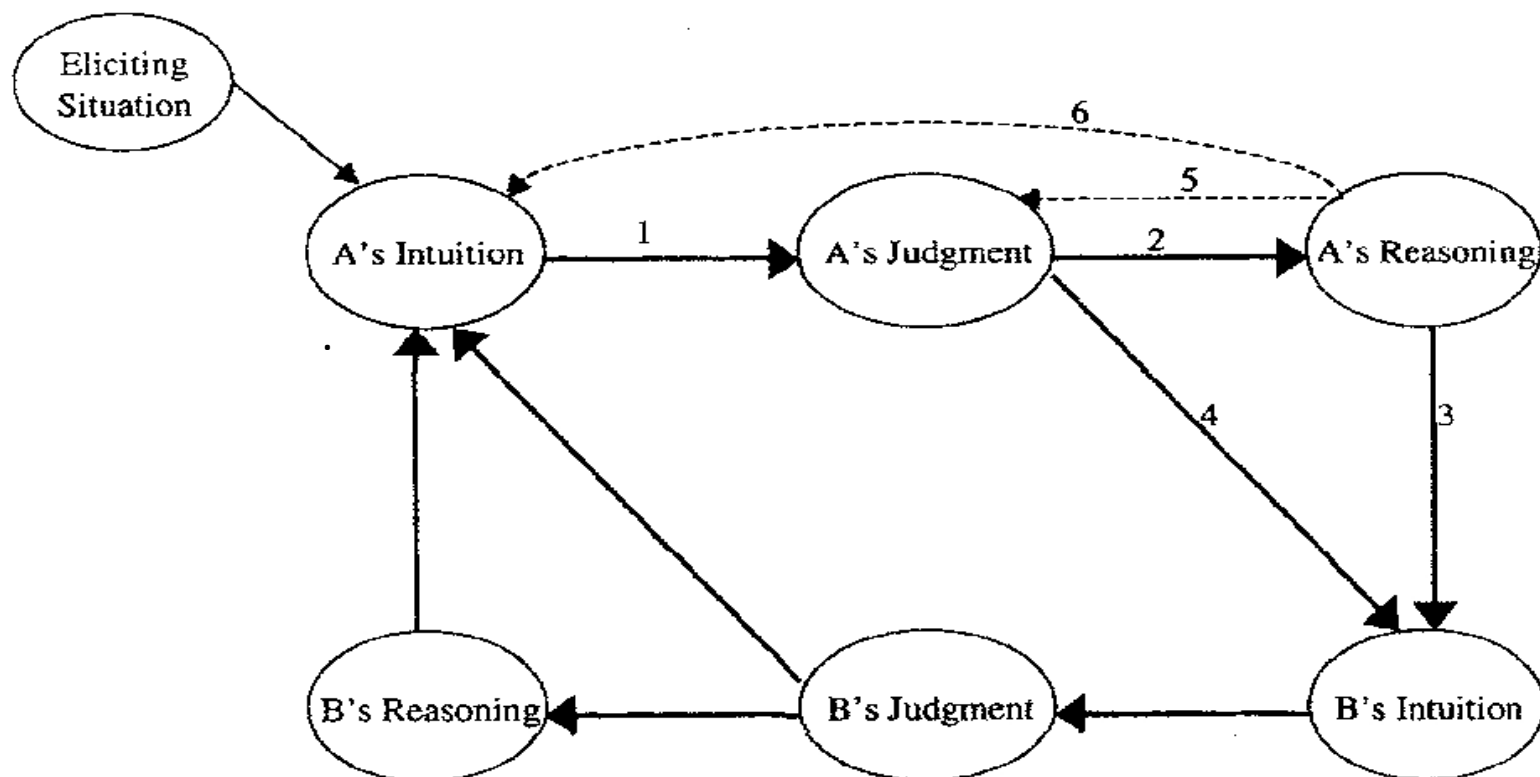
Figure 2. The social intuitionist model of moral judgment. The numbered links, drawn for Person A only, are (1) the intuitive judgment link (2) the post hoc reasoning link, (3) the reasoned persuasion link, and (4) the social persuasion link. Two additional links are hypothesized to occur less frequently: (5) the reasoned judgment link and (6) the private reflection link.

[A QUESTIONABLE DICHOTOMY?

- Haidt may be right about how many people reason their way *to* a moral judgment in cases like those he's concerned with (vs. cases where their intuitions are conflicting or weak).

- Instead, I'd question his dismissive treatment of "private" (= non-social) ex post facto moral reasoning as mere rationalization – at any rate, in competent moral reasoners.

- I think he works from too sharp a dichotomy between intuition (often, though not necessarily, emotion-based) and reasoning, as assigned to distinct cognitive processing systems [often labelled I and II].

- But some preliminary criticisms and clarifications are in order, before I get to my main argument.]

EVIDENCE FROM TABOO-CASES?

- Haidt's evidence for his model comes primarily from cases of harmless taboo-violation, e.g. brother-sister incest without harmful consequences, or using the national flag to clean the toilet.

- Subjects are said to be "dumbfounded" when their reasons for making a judgment of wrong are all shown to be inapplicable yet they persist (albeit uneasily) in making the judgment.

- [But does anyone think that *taboos* and our reactions to them are based on reasons and reasoning? I'll say more later about how these cases may be exceptional as regards norms of reasoning.]

- Dumbfounding was found to be more prevalent in lower social classes and non-Western societies.   [Perhaps Haidt's focus on taboo cases can be explained by a further point he's trying to establish: that morality covers, not just issues of harm and unfairness, but also issues of disgust and disrespect that are less stressed among educated inhabitants of Western societies.]

A HUMEAN APPROACH?

- **Haidt claims Hume as his philosophic hero – against Kantian views in psychology, put forth (most notably) by Piaget, Kohlberg, and Turiel.**

- **He often quotes Hume on reason as slave of the passions – and speaks approvingly of Hume's analogy between moral and aesthetic judgment.**

  - **[However, like some other psychologists claiming to follow Hume, he seems unaware that Hume's moral sentiments aren't just "gut feelings." Instead, they've been corrected to eliminate biases resulting from our tendency to sympathize more with others closer to us.**

  - **Hume's rather narrow notion of "reason" doesn't apply, but the correction involves a (positive) kind of "rationalization" of our passions, for the sake of *consistent* moral judgment.]**

[THE ROLE OF REASON

- Haidt isn't denying that emotions are tied to an agent's *reasons*. For that matter, he sees his results as illuminating "human" rationality, not as implying that people aren't rational. It's "reason" in the sense of reason*ing* that he debunks, at any rate in private moral thought.

- But I'll eventually suggest that the first two links in Haidt's causal model, including intuition, can be seen as together constituting a passage of reasoning, based on a variant of abduction – what's usually called "inference to the best explanation."

- Though the second step is ex post facto, it can still have causal significance – for *later* intuition and judgment – and it's normally required for *competent* moral judgment.]

THE AESTHETICS ANALOGY

- Since Haidt claims to favor Hume's analogy between moral and aesthetic judgment, let's consider aesthetic judgment.

- If we're really forming an aesthetic *judgment,* as needed for the analogy, we can't just stop with a report of our immediate likes or dislikes (as Haidt suggests for morality – and tells me is how he judges art works).

- We must in some way evaluate the object we like or dislike –as *worthy* of our own *and others'* aesthetic regard, say – so that our reaction to it isn't just personal or idiosyncratic.

- In fact, Hume's own account of aesthetic judgment is essentially an "ideal observer" account that rests on explaining the features of a person in the appropriate position to judge the worth of an aesthetic object.

- In our everyday judgments, we typically go on to cite some specific feature of the object that tends to support our judgment – yes, with a bias toward confirmation, in the sense that support is what we're looking for, but we needn't therefore be impervious to disconfirmation.

A CASE FROM ORDINARY THOUGHT

- Suppose I'm in a museum, attracted to a seascape, with an inclination to rate it above scenes of mountains or fields.

- But I'm aware that I tend to like water scenes generally. They remind me of enjoyable childhood vacations at the shore.

- So in defense of my immediate judgment, I note a specific feature that explains my reaction: that the artist's rapid brushwork conveys the motion of shimmering light.

- One needn't be a philosopher, or all that "cognitive" about art, to get this far – even if lots of people don't.

POSSIBLE LONGER-TERM EFFECTS

- So what might happen in the painting case, if we look beyond a single instance of judgment, as depicted in Haidt's model?

    - Having identified the effective use of rapid brushwork as a reason I have for a high opinion of the painting, I might be led to look for that feature in other paintings.

    - If I find it in paintings other than seascapes, I might also judge them good, whether or not I like them as much – though perhaps I'll eventually come to like them more than I do now.

    - Maybe I'll also give a lower aesthetic rating to seascapes I like that I can find no independent reason for judging good. Eventually, I might even come to like them less.

- What we have here is a feedback loop: the emotional dog *chasing* its rational tail.

STEPS OF REASONING

- Even before the stage of post hoc reasoning, the causal link between intuition and judgment might be seen as an immediate inference.

- It's an instance of *abduction*, or what's now mainly called "inference to the best explanation" – though C.S. Peirce, who invented the term, didn't specify "best." In the painting case I'm explaining my instant reaction to the seascape by its aesthetic value.

- That's not enough for reasoning in Haidt's sense, since it's effortless and automatic in most cases, involving only a single conscious step.

- But going on to give reasons to justify the judgment can be seen as a conscious abduction – at least if the notion is extended beyond causal or scientific explanation.

A MORAL ANALOGUE

- Now let's turn to a roughly parallel case of *moral* judgment.

- Suppose I recoil at someone's recent proposal to use genetic engineering to improve people morally (lest purely cognitive improvements produce evil madmen).

- Though I immediately judge it wrong, I'm aware that I might just be reacting to a reminder of the Nazi eugenics program. So I attempt to spell out an independent objection to it.

- I note that carrying out the proposal would mean empowering some subgroup to decide what traits to engineer, and that even a benign subgroup with such power (Aristotelians rather than Nazis) would be worrisome.

A DISANALOGY

- Of course, there are important differences in the way we treat moral vs. aesthetic judgments.

- For one thing, we're more likely to defer to experts in aesthetics, whereas at least in the modern era the assumption is that we're all competent to make moral judgments.

- No one says: "I don't know much about morality, but I know what I disapprove of" (or "...what disgusts me").

- But as noted earlier, the analogy pertains only to cases where we do go beyond a report of our feelings to venture an aesthetic *judgment*, meant to say that others in an appropriate position would share the reaction.

- What we leave to experts may be the task of justifying the judgment adequately.

MINIMAL MORAL COMPETENCE

- What we mean, or should mean, by the assumption of universal moral competence is that every normal human has the *capacity* to make sound moral judgments, not that everyone exercises it, or even develops a reliable disposition to exercise it.

- As Haidt maintains, people do need social input to develop the capacity, and often to raise questions of the sort that would prompt them to exercise it by looking for reasons.

- But it's sometimes reasonable to make a moral (or other) judgment even though one can't justify it adequately, as in Haidt's "dumbfounding" cases. Here, too, the job of justification may be left to experts – whether experts on moral reasoning or those with more developed moral sensitivity.

INCOMPETENT MORAL JUDGMENT

- In an experiment meant to show the moral relevance of disgust, Haidt and Wheatley found that subjects who'd been hypnotized to feel disgust at morally innocuous words came out with harsher condemnations of wrongdoers in passages containing those words.

- On the other hand, only a minority of subjects condemned the agent in a case of clearly permissible action that was included for comparison: a student council president choosing topics for discussion, with the aim of appealing to both faculty and students.

- One subject who did judge that case wrong is quoted as citing obviously inadequate reasons: that the passage reminded him of his bad times in high school, and the president seemed to be a "popularity-seeking snob."

- But stopping with these personal reactions wouldn't satisfy minimal standards of competent moral judgment.

NOT SO INCOMPETENT

- By contrast, Haidt's dumbfounded subjects did acknowledge the inadequacy of their reasons, at least with some prodding by the experimenter (which was absent from the post-hypnotic case).

- Their refusal to give up a judgment when at a loss for reasons needn't be taken as incompetent, though.

  - Think of a detective who senses that something is "wrong" (= amiss) in a crime scene but can't say what. Rather than dropping the judgment, he should continue *looking* for reasons to support it – or perhaps just wait till they occur to him, or until someone or something else provides them.

  - The task of spelling out adequate reasons for ruling out incest in a case that's stipulated to be harmless may reasonably be left to philosophers and other "cognizers" about morality (e.g. those familiar with rule-utilitarianism) or to others with more developed moral sensitivity.

REASONS NOT TO REASON?

- In fact, where taboos concern matters of general moral importance, ordinary agents – or even experts under ordinary circumstances – can be seen as having a second-order ("exclusionary") reason *not* to engage in reasoning about possible exceptions in particular cases.

  - Railton has argued that the subjects dumbfounded by Haidt's incest case may be emotionally attuned to reasons for avoiding a *risk* of serious harms (especially for frivolous benefits, such as those suggested in the passage describing the case).

  - Like the detective, they may be responding to reasons they can't articulate – though unlike the detective, maybe they shouldn't try to assess those reasons clearly, lest they be tempted to make risky exceptions.

- Of course, there's lots of room for error in everyday contexts about whether a particular moral proscription *should* be treated as a taboo – and taboos should be open to question in more removed contexts.

TO SUM UP

- **In ordinary cases, and for ordinary thinkers, moral competence does seem to involve readiness to engage in a certain modicum of reasoning.**

  - **Haidt's link between intuition and judgment amounts to an immediate abductive inference.**

  - **The ex post facto identification of reasons confirming the judgment can be seen as a further abductive inference completing a conscious two-step process that would satisfy Haidt's definition of reasoning.**

- **One might think of post hoc reasoning as analogous to "follow-through" on a golf-swing: though it comes only after one hits the ball, without it one hasn't hit the ball properly.**

RAWLSIAN INTUITIONS

- Singer takes the empirical evidence he cites to undermine Rawls's method of reflective equilibrium (henceforth r.e.) insofar as the method relies on intuitions about particular cases.

- However, Rawls specifies "considered" intuitions as the input to r.e.   These typically aren't what the empirical work is about, as illustrated by Haidt's definition.

- In his *Theory of Justice* (*TOJ*) Rawls explains the notion of a "considered" intuition simply as meant to rule out intuitions based on distortive emotional influence.

- But his earlier "Outline of a Decision Procedure for Ethics" gives a fuller account, in its conditions on the considered judgments of competent moral judges, as input to Rawls's decision procedure.

RAWLS'S DECISION PROCEDURE

- When Rawls requires in "Outline" that considered judgments be "intuitive" (p. 183), he explains this as meaning just that they're not consciously and systematically derived from general moral principles.

- But they might still be based on factual inquiry, reflection on consequences in relation to alternatives, and/or application of a common sense rule, as forms of pre-theoretical reflection.

- Rawls's method represents a strand of thought in contemporary metaethics that attempts to do without metaphysical commitment to real (= mind-independent) moral properties, but without denying the objectivity of ethics.

- Instead Rawls thinks that what makes ethics objective is the possibility of validating moral judgments by a decision procedure – in the sense of a reliable method for validating moral principles.

FOUR STEPS

1. Define a class of competent moral judges by their normal intelligence and relevant factual knowledge, plus possession of standard "intellectual virtues" (reasonableness – including open-mindedness; cf. Haidt's subjects) and a "virtue of moral insight" (sympathetic knowledge of affected interests), pp. 178ff).

2. Elicit the judges' considered intuitive judgments on familiar, relatively easy real-life cases. (For conditions on such judgments, including "stability," see pp. 181ff.)

3. "Explicate" the judgments with general moral principles from which they (or most of them) can be derived.

4. Test the principles, e.g. by deriving judgments on new and harder cases, showing that the principles can sometimes override intuitions.

ETHICS VS. SCIENCE

- It's also questionable whether Rawls's method takes intuitions as evidence of moral truth, by analogy to science, as Singer and others seem to assume.

- Rawls makes it clear in later work on metaethics that he eschews truth in favor of a distinct notion of "objectivity." In "Outline" he characterizes objectivity just in terms of a "reliable method of validation" when he mentions the analogy to science (see p. 177)

- Assuming that science does discover truth (as Singer, Haidt, and others – but not all philosophers of science – seem to do), Rawls's analogy between ethics and science must be limited.

- In fact, what he says is that the appropriate analogue to his method is *inductive logic* – not scientific reasoning itself, but the construction of a normative system for determining when scientific reasoning is valid.

BACK-AND-FORTH REASONING

- Singer also seems to take Rawls's method as one-way: formulating principles to "match" intuitions on cases – the opposite of Singer's "top-down" ultilitarian approach, which starts with an intuitively appealing general principle (the Principle of Utility) and derives particular case-judgments from it.

- "Outline" may at first give that impression, though ultimately Rawls allows for modifying certain intuitions to conform to the principles that are formulated to "explicate" others. (See pp. 188f., where he requires that the principles be able to "hold their own" against conflicting intuitions.)

- In Rawls's later *TOJ* the mutual adjustment of principles and intuitions is presented as central to his decision method, which at that point first gets the name "reflective equilibrium."

- Footnotes in *TOJ* (p. 18, n. 7, and p. 507, n. 34) indicate that he sees his eventual account there as jibing with work on inductive logic by Goodman and Quine that appeared after "Outline."

CONTRASTING METHODS

- It might be useful to sum up three different theoretical methods, in ethics and elsewhere:

    1. "Top-down": particular judgments deduced from general principles (e.g., Singer's utilitarianism)

    2. "Bottom-up": general principles derived by induction from particular judgments (e.g., observation in empirical sciences)

    3. "Back-and-forth": mutual adjustment of case-intuitions and theoretical principles (Rawls's r.e., taken as analogous to Goodman's way of deriving principles of inductive logic)

- Singer seems to treat Rawls's view as an instance of 2, whereas it really belongs under 3.

"STABLE" INTUITIONS

- Another important feature of Rawls's account in "Outline" that we should note for future reference is his condition (vi) of "stability" on considered intuitive judgments (pp. 192f.).

- He writes:

   It is required that the judgment be stable, that is, that there be evidence that at other times and at other places competent judges have rendered the same judgment on similar cases, understanding similar cases to be those in which the relevant facts and the competing interests are similar. The stability must hold, by and large, over the class of competent judges and *over their judgments at different times*.... (p. 182; italics mine)

- This apparently would rule out case-intuitions that change when the relevant cases are presented in a different order (as in Wiegmann's study of "order effects," which we'll read later). They count as input into his system only if their "considered" versions are stable.

GREENE VS. DEONTOLOGY

- Greene, a neuroscientist with a Ph.D. in philosophy, gives empirical evidence from neuroimaging that deontological intuitions are "driven by" emotion.

- In philosophy deontological ethics emphasizes moral rules and is focused on rights and duties, whereas consequentialism considers only consequences.

- However, Greene deals with characteristic *judgments* rather than philosophical theories. A characteristic deontological judgment would be that killing one innocent person is wrong even if it's needed to save several others.

- To avoid the objection that deontological judgments by definition are based on respect for moral rules rather than emotions (p. 37), Greene understands the relevant notions as referring to "*psychological natural kinds*": instinctive ways of thinking that philosophers identified but didn't invent.

- Following Kant, deontologists tend to avoid reliance on emotion. Hence Greene's title (from Nietzsche).

"COGNITION" VS. "EMOTION"

- 'Cognition' in a broad sense (= information processing) also applies to emotion, so Greene uses quotation marks for his narrower sense [which basically assigns emotion to System I and "cognition" to System II], treating "cognitive" behavioral responses as more flexible than emotional responses in light of different circumstances.

- Emotional and "cognitive" processes tend to be associated with different parts of the brain.

- Though Greene favors the Humean view that *all* moral judgment has an emotional component, he takes a different ("alarmlike") kind of emotion to be involved in deontological judgments. [So he really should use quotation marks for "emotional" too!]

- His view departs from both extremes in psychology (e.g., Kohlberg vs. Haidt, seen as taking all moral judgments as "cognitive" or "emotional," respectively) — and also from historical views in philosophy, which associate cognition with deontology (Kant) and emotion with consequentialism (Hume, Adam Smith).

RESEARCH ON THE TROLLEY PROBLEM

- **Greene applied neuroimaging to a version of the trolley problem, noting that people favor a characteristically consequentialist judgment in the standard trolley case and a deontological judgment in the footbridge case.**

- **He and his fellow researchers hypothesized that what makes the difference is the type and thus the intensity of emotion associated with "(up close and) personal" vs. "impersonal" harm: pushing the man off the footbridge vs. merely hitting a switch.**

- **Evolution yields a plausible rationale for an innate "alarmlike" response to personal violence, whereas impersonal cases allow for detached calculation.**

- **Greene's brain studies confirmed that contemplation of personal moral dilemmas produced relatively greater activity in emotion-related areas.**

- **They also confirmed his prediction that it would take longer to override one's natural emotional response and *permit* action in the footbridge case than it would to prohibit action on the basis of emotion – whereas no corresponding difference would be found in the switch case.**

GREENE'S CLAIMS

- Greene explains deontological responses on the basis of evolution (pp. 366f.) — and deontological moral theory (referring, e.g., to rights) as a post hoc rationalization: a "confabulation" of reasons for responses whose innate emotional bases we don't understand.

- He sums up deontological responses (p. 63) as a "'nay-saying' voice" that "can be overridden…, but as far as the voice itself is concerned, [leaves] no room for negotiation."

- By contrast, consequentialism is more "cognitive" because it's systematic and aggregative – a weighing process that "makes nearly everything negotiable" (p. 64). In that sense, it may be said to involve "currency" rather than "alarm" emotions.

- Consequentialist representations are "neutral" in not automatically giving rise to behavioral dispositions, so that the agent can combine them without being pulled in different directions. This is how they allow for more flexible behavioral response.

GREENE'S EVENTUAL ARGUMENT

- Greene gives a more cautious version of his conclusion toward the end of his piece: that the burden is on deontologists to explain how our emotions could be responding to the principles they appeal to. In the first instance, he's attacking *rationalist* deontologists, but he claims (p. 75) that his argument applies to others as well.

- In its eventual form (pp. 67ff.; see esp. p. 69.) his argument does depend on the evolutionary origins of deontological responses – but with particular reference to the fact that certain morally irrelevant features of the evolutionary environment warranted "alarm" emotions. This *wouldn't* apply to "characteristically consequentialist" responses, which on his account are motivated solely by "currency" emotions.

- His assumption, following Singer, is that distance is morally irrelevant. Also, what evolution favored was "biological fitness," i.e. gene-replication, rather than individual or social interests. [Of course, behavioral flexibility may also promote biological fitness, so in that sense consequentialist responses are also products of evolution, contra Singer.]

ASYMMETRICAL ORDER EFFECTS

- **Wiegmann describes an experiment showing that intuitive judgments of permissibility are affected by order of presentation when and only when a relevantly similar case judged impermissible is presented first.**

  - *The experiment used variants of the trolley case, labeled "Standard" (flipping a switch) and "Push" (the footbridge case), along with several intermediate options to heighten perceived similarity from case to case (see p. 9).*

  - *Presenting Push first made subjects' more likely to judge Standard impermissible. But presenting Standard first didn't make them any more likely to judge Push permissible.*

- **In short, these results exhibited a pattern of asymmetric "judgment transfer": negative judgments on cases presented first affected judgments on cases that would have been judged positively in isolation, but not vice versa.**

POSSIBLE EXPLANATIONS

- **Wiegman suggests two possible explanations, reflective and emotional, of the asymmetry of order effects:**

  - **Since prohibitions call for justification in a way that permissions don't, subjects presented with Push first might have thought of a principle justifying their prohibition which they then applied to later cases – whereas when Standard was presented first, they didn't bother justifying their permission.**

  - **Alternatively, since emotions can affect reflective reasoning, but not vice versa, the results might have been due to differential emotional engagement (cf. Greene), whereby the "personal" features of Push evoke transferrable hard-wired moral emotions in a way that Standard does not.**

- **Combinations are also possible, e.g. if the strength of our emotional response makes Push, but not Standard, capable of transferring justifying principles to later cases but immune to transfer itself.**

RELEVANCE TO INTUITIONS

- **Wiegmann takes his findings as evidence that moral intuitions can be influenced by a morally irrelevant factor (= order of presentation).**

- **He ends by weighing the naturalness vs. artificiality of the experimental settings producing his results, to see whether it's appropriate to generalize them to real-life cases.**

  - **Wiegmann notes that, though the experiments involved laymen rather than philosophers, it's questionable whether philosophers have better intuitions than ordinary folk (= "the expertise defense").**

  - **Though natural enough for people interested in moral questions, especially philosophers, the experimental settings required giving an answer right away, without the opportunity to (re)consider cases in light of others presented later. Wiegmann therefore says it would be "over-hasty" to conclude that intuitions can't be used as evidence (p. 20; cf. Singer).**

[ALTERNATIVES

- Thomson, after a long time considering the problem she raised, eventually concluded that pulling the switch in the Standard case was *not* permissible.

- Also, while Wiegmann's order effects may be irrelevant to moral properties or facts, his explanations might suggest that they're relevant to competent moral *reasoning and/or judgment.*

  - Note that he also includes as irrelevant such features of the subject as education (p. 17), along with features of the case description such as emphasis on the brutality of the killing.

  - Perhaps our sense that switching the trolley in the Standard case is permissible also has a psychological explanation that calls it into question. Considering the Standard case from a first-person perspective, we're likely to think first of the need to avoid killing the five ("Swerve!"), only then noting that we'll be killing one instead.

  - It would be useful to test "within scenario" effects by comparing a version of the Standard case that starts by describing the killing with a reasonably accurate depiction of the "brutality" involved, repeated in both versions.]

PHILOSOPHERS' INTUITIONS

- Tobia, Buckwalter, and Stich (TBS) ran experiments on another "framing effect," the actor/observer bias, to test the expertise defense and found that, while both philosophers and nonphilosophers exhibited the bias, they exhibited it in reverse directions:

  - In response to the Standard trolley case, nonphilosophers mainly judged it permissible for a third person to pull the switch, but not for themselves to do so.

  - By contrast, philosophers mainly judged it permissible for themselves to pull the switch, but not for a third person.

- They don't consider possible explanations, though, as Wiegmann does for his results. What would you suggest?

POSSIBLE EXPLANATIONS

- TBS also found that in another case (Williams's case of Jim and the natives) philosophers were more likely to impose an obligation on themselves than on Jim, whereas nonphilosophers exhibited the reverse actor/observer bias.

- It's been suggested that philosophers may be influenced by a tendency to view themselves as heroes, whereas nonphilosophers anticipate guilt for the killing.

- Note that this explanation might be seen as making out the first-person perspective as *distortive* – in line with Arpaly's approach.

- Another possibility that occurred to me is that imposing an obligation on someone and allowing him to do something are *acts* of a sort and might well be evaluated differently terms when directed toward oneself vs. others. You're being hard on someone (perhaps unfairly) when you require more of him – or you're letting him off the hook if you require less.