# F/W—Util—Blocked

## NC

### NC Util—vs means

#### The standard is maximizing happiness.

#### Revisionary intuitionism is true and justifies util

Yudkowsky 8 Eliezer Yudkowsky (research fellow of the Machine Intelligence Research Institute; he also writes Harry Potter fan fiction). “The ‘Intuitions’ Behind ‘Utilitarianism.’” 28 January 2008. LessWrong. http://lesswrong.com/lw/n9/the\_intuitions\_behind\_utilitarianism/

I haven't said much about metaethics - the nature of morality - because that has a forward dependency on a discussion of the Mind Projection Fallacy that I haven't gotten to yet. I used to be very confused about metaethics. After my confusion finally cleared up, I did a postmortem on my previous thoughts. I found that my object-level moral reasoning had been valuable and my **meta-level moral reasoning had been worse than useless.** And this appears to be a general syndrome - **people do** not do **much better when discussing whether torture is** good or **bad than when they discuss the meaning of "good" and "bad". Thus, I deem it prudent to keep moral discussions on the object level** wherever I possibly can. Occasionally **people object** to any discussion of morality on the grounds **that morality doesn't exist**, and in lieu of jumping over the forward dependency to explain that **"exist" is not the right term to use** here, I generally say, "But **what do you do anyway?**" and **take the discussion back down to the object level.** Paul Gowder, though, has pointed out that both the idea of choosing a googolplex dust specks in a googolplex eyes over 50 years of torture for one person, and the idea of "utilitarianism", depend on "intuition". He says I've argued that the two are not compatible, but charges me with failing to argue for the utilitarian intuitions that I appeal to. Now "intuition" is not how I would describe the computations that underlie human morality and distinguish us, as moralists, from an ideal philosopher of perfect emptiness and/or a rock. But I am okay with using the word "intuition" as a term of art, bearing in mind that "intuition" in this sense is not to be contrasted to reason, but is, rather, the cognitive building block out of which both long verbal arguments and fast perceptual arguments are constructed. **I see** the project of **morality as a project of renormalizing intuition.** We have intuitions about things that seem desirable or undesirable, intuitions about actions that are right or wrong, intuitions about how to resolve conflicting intuitions, intuitions about how to systematize specific intuitions into general principles. **Delete all** the **intuitions, and** you aren't left with an ideal philosopher of perfect emptiness, **you're left with a rock. Keep all your** specific **intuitions and** refuse to build upon the reflective ones, and you aren't left with an ideal philosopher of perfect spontaneity and genuineness, **you're left with a** grunting **caveperson** running in circles, due to cyclical preferences and similar inconsistencies. "Intuition", as a term of art, is not a curse word when it comes to morality - there is nothing else to argue from. **Even modus ponens is an "intuition"** in this sense - **it**'s **just** that modus ponens **still seems like a good idea after being** formalized, **reflected on**, extrapolated out to see if it has sensible consequences, etcetera. So that is "intuition". However, Gowder did not say what he meant by "utilitarianism". Does utilitarianism say... That right actions are strictly determined by good consequences? That praiseworthy actions depend on justifiable expectations of good consequences? That probabilities of consequences should normatively be discounted by their probability, so that a 50% probability of something bad should weigh exactly half as much in our tradeoffs? That virtuous actions always correspond to maximizing expected utility under some utility function? That two harmful events are worse than one? That two independent occurrences of a harm (not to the same person, not interacting with each other) are exactly twice as bad as one? That for any two harms A and B, with A much worse than B, there exists some tiny probability such that gambling on this probability of A is preferable to a certainty of B? If you say that I advocate something, or that my argument depends on something, and that it is wrong, do please specify what this thingy is... anyway, I accept 3, 5, 6, and 7, but not 4; I am not sure about the phrasing of 1; and 2 is true, I guess, but phrased in a rather solipsistic and selfish fashion: you should not worry about being praiseworthy. Now, what are the "intuitions" upon which my "utilitarianism" depends? This is a deepish sort of topic, but I'll take a quick stab at it. First of all, it's not just that someone presented me with a list of statements like those above, and I decided which ones sounded "intuitive". Among other things, **if you try to violate** "**util**itarianism", **you run into paradoxes, contradictions**, circular preferences, **and other** things that aren't **symptoms of** moral wrongness so much as **moral incoherence.** After you think about moral problems for a while, and also find new truths about the world, and even discover disturbing facts about how you yourself work, you often end up with different moral opinions than when you started out. This does not quite define moral progress, but it is how we experience moral progress. As part of my experienced moral progress, I've drawn a conceptual separation between questions of type Where should we go? and questions of type How should we get there? (Could that be what Gowder means by saying I'm "utilitarian"?) The question of where a road goes - where it leads - you can answer by traveling the road and finding out. If you have a false belief about where the road leads, this falsity can be destroyed by the truth in a very direct and straightforward manner. When it comes to wanting to go to a particular place, this want is not entirely immune from the destructive powers of truth. You could go there and find that you regret it afterward (which does not define moral error, but is how we experience moral error). But, even so, wanting to be in a particular place seems worth distinguishing from wanting to take a particular road to a particular place. Our intuitions about where to go are arguable enough, but our intuitions about how to get there are frankly messed up. **After** the two hundred and eighty-seventh **research** study **showing that people will chop their own feet off if you frame the problem the wrong way, you start to distrust first impressions. When you've read enough research on scope insensitivity** - people will pay only 28% more to protect all 57 wilderness areas in Ontario than one area, **people will pay the same amount to save 50,000 lives as 5,000 lives**... that sort of thing... Well, the worst case of scope insensitivity I've ever heard of was described here by Slovic: Other recent research shows similar results. Two Israeli psychologists asked people to contribute to a costly life-saving treatment. They could offer that contribution to a group of eight sick children, or to an individual child selected from the group. The target amount needed to save the child (or children) was the same in both cases. Contributions to individual group members far outweighed the contributions to the entire group. There's other research along similar lines, but I'm just presenting one example, 'cause, y'know, eight examples would probably have less impact. If you know the general experimental paradigm, then the reason for the above behavior is pretty obvious - focusing your attention on a single child creates more emotional arousal than trying to distribute attention around eight children simultaneously. So people are willing to pay more to help one child than to help eight. Now, **you could** look at this intuition, and **think it was** revealing **some** kind of **incredibly deep moral truth** which shows that one child's good fortune is somehow devalued by the other children's good fortune. But what about the billions of other children in the world? Why isn't it a bad idea to help this one child, when that causes the value of all the other children to go down? How can it be significantly better to have 1,329,342,410 happy children than 1,329,342,409, but then somewhat worse to have seven more at 1,329,342,417? **Or you could** look at that and **say: "The intuition is wrong: the brain can't** successfully **multiply** by eight and get a larger quantity than it started with. **But it ought to**, normatively speaking." And once you realize that the brain can't multiply by eight, then the other cases of scope neglect stop seeming to reveal some fundamental truth about 50,000 lives being worth just the same effort as 5,000 lives, or whatever. You don't get the impression you're looking at the revelation of a deep moral truth about nonagglomerative utilities. It's just that the brain doesn't goddamn multiply. Quantities get thrown out the window. If you have $100 to spend, and you spend $20 each on each of 5 efforts to save 5,000 lives, you will do worse than if you spend $100 on a single effort to save 50,000 lives. Likewise if such choices are made by 10 different people, rather than the same person. As soon as you start believing that it is better to save 50,000 lives than 25,000 lives, that simple preference of final destinations has implications for the choice of paths, when you consider five different events that save 5,000 lives. (It is a general principle that Bayesians see no difference between the long-run answer and the short-run answer; you never get two different answers from computing the same question two different ways. But the long run is a helpful intuition pump, so I am talking about it anyway.) The aggregative valuation strategy of "shut up and multiply" arises from the simple preference to have more of something - to save as many lives as possible - when you have to describe general principles for choosing more than once, acting more than once, planning at more than one time. Aggregation also arises from claiming that the local choice to save one life doesn't depend on how many lives already exist, far away on the other side of the planet, or far away on the other side of the universe. Three lives are one and one and one. No matter how many billions are doing better, or doing worse. 3 = 1 + 1 + 1, no matter what other quantities you add to both sides of the equation. And if you add another life you get 4 = 1 + 1 + 1 + 1. That's aggregation. **When you've read enough** heuristics and **biases research, and enough coherence** and uniqueness **proofs for** Bayesian probabilities and **expected utility**, and you've seen the "Dutch book" and "money pump" effects that penalize trying to handle uncertain outcomes any other way, **then you don't see** the **preference reversals** in the Allais Paradox **as** revealing **some** incredibly **deep moral truth** about the intrinsic value of certainty. **It just goes to show that the brain doesn't** goddamn **multiply.** The primitive, perceptual intuitions that make a choice "feel good" don't handle probabilistic pathways through time very skillfully, especially when the probabilities have been expressed symbolically rather than experienced as a frequency. So you reflect, devise more trustworthy logics, and think it through in words. When you see people insisting that no amount of money whatsoever is worth a single human life, and then driving an extra mile to save $10; or when you see people insisting that no amount of money is worth a decrement of health, and then choosing the cheapest health insurance available; then you don't think that their protestations reveal some deep truth about incommensurable utilities. Part of it, clearly, is that **primitive intuitions don't successfully diminish the emotional impact of** symbols standing for **small quantities** - anything you talk about seems like "an amount worth considering". And part of it has to do with preferring unconditional social rules to conditional social rules. Conditional rules seem weaker, seem more subject to manipulation. If there's any loophole that lets the government legally commit torture, then the government will drive a truck through that loophole. So it seems like there should be an unconditional social injunction against preferring money to life, and no "but" following it. Not even "but a thousand dollars isn't worth a 0.0000000001% probability of saving a life". Though the latter choice, of course, is revealed every time we sneeze without calling a doctor. The rhetoric of sacredness gets bonus points for seeming to express an unlimited commitment, an unconditional refusal that signals trustworthiness and refusal to compromise. So you conclude that moral rhetoric espouses qualitative distinctions, because espousing a quantitative tradeoff would sound like you were plotting to defect. On such occasions, people vigorously want to throw quantities out the window, and they get upset if you try to bring quantities back in, because quantities sound like conditions that would weaken the rule. But you don't conclude that there are actually two tiers of utility with lexical ordering. You don't conclude that there is actually an infinitely sharp moral gradient, some atom that moves a Planck distance (in our continuous physical universe) and sends a utility from 0 to infinity. You don't conclude that utilities must be expressed using hyper-real numbers. Because the lower tier would simply vanish in any equation. It would never be worth the tiniest effort to recalculate for it. All decisions would be determined by the upper tier, and all thought spent thinking about the upper tier only, if the upper tier genuinely had lexical priority. As Peter Norvig once pointed out, if Asimov's robots had strict priority for the First Law of Robotics ("A robot shall not harm a human being, nor through inaction allow a human being to come to harm") then no robot's behavior would ever show any sign of the other two Laws; there would always be some tiny First Law factor that would be sufficient to determine the decision. Whatever value is worth thinking about at all, must be worth trading off against all other values worth thinking about, because thought itself is a limited resource that must be traded off. When you reveal a value, you reveal a utility. I don't say that morality should always be simple. I've already said that the meaning of music is more than happiness alone, more than just a pleasure center lighting up. I would rather see music composed by people than by nonsentient machine learning algorithms, so that someone should have the joy of composition; I care about the journey, as well as the destination. And I am ready to hear if you tell me that the value of music is deeper, and involves more complications, than I realize - that the valuation of this one event is more complex than I know. But that's for one event. When it comes to multiplying by quantities and probabilities, complication is to be avoided - at least if you care more about the destination than the journey. **When you've reflected** on enough intuitions, **and corrected enough absurdities, you** start to **see a common denominator, a meta-principle** at work, **which one might phrase as "Shut up and multiply."** Where music is concerned, I care about the journey. When lives are at stake, I shut up and multiply. It is more important that lives be saved, than that we conform to any particular ritual in saving them. And the optimal path to that destination is governed by laws that are simple, because they are math. **And that's why I'm a utilitarian** - at least when I am doing something that is overwhelmingly more important than my own feelings about it - which is most of the time, because there are not many utilitarians, and many things left undone.

#### Unless we can empirically verify some knowledge of morality, we have no way of discerning which actions violate previously established side constraints until evaluating the end states those actions produce. We cannot know that killing someone is wrong until we assess the consequences of their death. Given that the very idea of harm is determined by end states, it follows that we are obligated to positively influence those end states.

#### Means-based standards cannot prescribe action insofar as they do not permit empirical evaluations of how much risk of a rule violation is sufficient to justify moral criticism. Given that empirical considerations advise courses of actions based on unique contexts and rational beings are instinctively inclined to follow the course of action that offers the highest probability of producing a desirable end, only ends-based standards are consistent with morality’s stated purpose of serving as a guide for action.

#### The act-omission distinction is nonsensical—no brightline for what constitutes an action

Persson[Ingmar Persson (Department of Philosophy, Lund University, Kungshuset), “Two act-omission paradoxes,” Proceedings of the Aristotelian Society, vol. 104 (2), pp. 147-162, 2004]

There are two ways in which the act-omission doctrine**,** which implies that it may be **[is] permissible to let people die** or be killed when it is **[but] wrong to kill them, gives rise to** a **paradox. First, it may be** that **when you let a victim be killed, you let yourself kill this victim**. On the assumption that, if it would be wrong of you to act in a certain fashion, **it would be** wrong of you [to] let yourself act in this fashion, this yields the paradox that it is **both permissible and impermissible to** let yourself **act in this fashion. Second, you may let yourself kill** somebody **by letting an action** **you have already initiated cause death**, e. g., **[i.e] by not lending** **a** helping **hand to somebody** **you have pushed**. This, too, yields [a contradiction] the paradox that it is both permissible and impermissible to let yourself kill if you are in a situation in which [when] killing is impermissible but letting be killed permissible.

### NC Util—No intent/foresight

#### The intent-foresight distinction can’t guard against absurd conclusions

**Di Nucci 14** writes[[1]](#footnote-1)

According to the problem of closeness, **the normative distinction between intending** harm **and** merely **foreseeing harm is unworkable because we can always argue that the relevant harm was merely foreseen** and the Doctrine of Double Effect offers no criterion to rule out any of these cases: so that (to take one of Philippa Foot’s famous scenarios) **if we blow up a fat guy** to pieces **whose body is obstructing the exit of the cave where we are stuck,** then **we can say that the death of the fat guy was a** merely **foreseen** (unintended) **consequence of freeing up the cave**’s mouth**– and the Doctrine has no criterion to stop this and** endless **other implausible applications.**

### NC Util—AT: constitutivism

#### Attempts to derive morality from principles constitutive of action or agency fail because there is no reason that a rational person would care about whether what they’re doing is an action or a schmaction.

Enoch 6 (David Enoch, Prof @ Hebrew University, Jerusalem, “Agency, Shmagency: Why Normativity Won’t Come from What Is Constitutive of Action,” 2006)//Miro

The problems this strategy was supposed to solve—the skeptical challenge, the externalist challenge, and the antinaturalist challenge— all revolve around the idea that the normative, though somehow essentially tied to us and our desires (or at least our will) can nevertheless not be as arbitrary as our (usual) motivations. But then how is the fact that some motivations are constitutive of action supposed to help? Consider Rosati first. According to Rosati, we are to think of a (perhaps ideally) autonomous agent stepping back from her desires (of whatever order) because she sees them—when deliberating and evaluating—as normatively arbitrary. And we are to think of her as troubled by the fact that it is hard to see what could possibly give the answers she is looking for because all facts of her psychology are just as arbitrary as her desires. She then finds out that some parts of her psychological makeup are unique in that they are such that without them she would not have qualified as an agent at all. Knowing that, is she supposed to be relieved? Why does it matter, as far as the question of normative arbitrariness is concerned, that some parts of her psychology have this necessary-for-agency status? Why shouldn’t our agent treat the motives and capacities constitutive of agency as normatively arbitrary? Why shouldn’t she treat the very fact that they are constitutive of agency as normatively arbitrary? She is, remember, stepping back from her desires, attempting a kind of detached scrutiny, evaluation, and choice. But then how is the constitutive-of-agency status at all relevant? What is it to her, so to speak, if some of her motives and capacities enjoy such a status? Or consider Korsgaard’s hope of grounding a reply to the skeptic in what is constitutive of action. We are to imagine, then, someone who remains indifferent when we tell him that his actions are immoral or irrational. He then reads Korsgaard and is convinced that self-constitution is a constitutive aim of action, so that you cannot even count as an agent and your bodily movements cannot even count as actions unless you aim at self-constitution of the kind Korsgaard has in mind. And assume that our skeptic is even convinced that—miraculously24—morality and indeed the whole of practical rationality can be extracted from the aim of self-constitution. Do we have any reason to believe that now he will care about the immorality or irrationality of his actions? Why isn’t he entitled to respond along the following lines: “Classify my bodily movements and indeed me as you like. Perhaps I cannot be classified as an agent without aiming to constitute myself. But why should I be an agent? Perhaps I can’t act without aiming at self-constitution, but why should I act? If your reasoning works, this just shows that I don’t care about agency and action. I am perfectly happy being a “shmagent”—a nonagent who is very similar to agents but who lacks the aim (constitutive of agency but not of “shmagency”) of self-constitution. I am perfectly happy performing “shmactions”—nonaction events that are very similar to actions but that lack the aim (constitutive of actions but not of shmactions) of selfconstitution.” Has Korsgaard put us in a better spot vis-à-vis this why be-an-agent (rather than a shmagent) problem than we were vis-à-vis the why-be-moral or why-be-rational challenges with which we—or at least Korsgaard—started? Consider again the example of the house and the shoddy builder, and suppose we manage to convince him that certain standards—standards he previously did not care about and regularly failed to measure up to—are constitutive of being a house. It seems he is entitled to respond: “Very well then, I guess I am not engaging in the project of building a house but rather in the project of building a “shmouse,” of which these standards aren’t constitutive. So what is it to me how you classify my project?” At times Korsgaard writes as if she thinks no such retort—either in the house case or in the metaethical or metanormative case—is possible. In Lewis’s (1996, 60) terms, at times Korsgaard writes as if she believes that the threat that your inner (and outer) states will fail to deserve folk-theoretical names (such as “action”) is indeed a threat that will strike terror into the hearts of the wicked.25 But no support is offered for this surprising claim. And notice that Korsgaard’s problem here is not merely that the skeptic is unlikely to be convinced by such a maneuver. The problem runs deeper than that because the skeptic should not be convinced.26 However strong or weak the reasons that apply to him and require that he be moral, surely they do not become stronger when he realizes that unless he complies with morality his bodily movements will not be adequately described as actions. Notice that the problem is not that action does not have a constitutive aim, or that there are no motives and capacities constitutive of agency. Indeed, I am here granting these claims for the sake of argument. Nor is the problem that such constitutive aims, motives, and capacities are philosophically uninteresting. For all I am about to say, they may be able to explain much that is philosophically important as well as interesting.27 The problem is just that it is hard to see how the constitutivist strategy can serve to ground normativity or to solve the metanormative problems it was supposed to solve.

### NC Util—AT: rule util

Devolves to act util.

1. “Maximize utility” is the utility maximizing rule.

2. If living wage is an exception to the rule, that proves that the rule “follow the aff principles except for living wage” is net better than the aff.

### NC Util—precludes

#### Epistemic modesty means that we must avoid existential risk—even with low probability of util, we still have to concede high leverage issues.

Bostrom 9 (Nick Bostrom , Professor, Faculty of Philosophy & Oxford Martin School, “Moral uncertainty – towards a solution?”, January 1 2009)//Miro

It seems people are overconfident about their moral beliefs. But how should one reason and act if one acknowledges that one is uncertain about morality – not just applied ethics but fundamental moral issues? if you don't know which moral theory is correct? It doesn't seem you can simply plug your uncertainty into expected utility decision theory and crank the wheel; because many moral theories state that you should not always maximize expected utility. Even if we limit consideration to consequentialist theories, it still is hard to see how to combine them in the standard decision theoretic framework. For example, suppose you give X% probability to total utilitarianism and (100-X)% to average utilitarianism. Now an action might add 5 utils to total happiness and decrease average happiness by 2 utils. (This could happen, e.g. if you create a new happy person that is less happy than the people who already existed.) Now what do you do, for different values of X? The problem gets even more complicated if we consider not only consequentialist theories but also deontological theories, contractarian theories, virtue ethics, etc. We might even throw various meta-ethical theories into the stew: error theory, relativism, etc. I'm working on a paper on this together with my colleague Toby Ord. We have some arguments against a few possible "solutions" that we think don't work. On the positive side we have some tricks that work for a few special cases. But beyond that, the best we have managed so far is a kind of metaphor, which we don't think is literally and exactly correct, and it is a bit under-determined, but it seems to get things roughly right and it might point in the right direction: The Parliamentary Model. Suppose that you have a set of mutually exclusive moral theories, and that you assign each of these some probability. Now imagine that each of these theories gets to send some number of delegates to The Parliament. The number of delegates each theory gets to send is proportional to the probability of the theory. Then the delegates bargain with one another for support on various issues; and the Parliament reaches a decision by the delegates voting. What you should do is act according to the decisions of this imaginary Parliament. (Actually, we use an extra trick here: we imagine that the delegates act as if the Parliament's decision were a stochastic variable such that the probability of the Parliament taking action A is proportional to the fraction of votes for A. This has the effect of eliminating the artificial 50% threshold that otherwise gives a majority bloc absolute power. Yet – unbeknownst to the delegates – the Parliament always takes whatever action got the most votes: this way we avoid paying the cost of the randomization!) The idea here is that moral theories get more influence the more probable they are; yet even a relatively weak theory can still get its way on some issues that the theory think are extremely important by sacrificing its influence on other issues that other theories deem more important. For example, suppose you assign 10% probability to total utilitarianism and 90% to moral egoism (just to illustrate the principle). Then the Parliament would mostly take actions that maximize egoistic satisfaction; however it would make some concessions to utilitarianism on issues that utilitarianism thinks is especially important. In this example, the person might donate some portion of their income to existential risks research and otherwise live completely selfishly. I think there might be wisdom in this model. It avoids the dangerous and unstable extremism that would result from letting one’s current favorite moral theory completely dictate action, while still allowing the aggressive pursuit of some non-commonsensical high-leverage strategies so long as they don’t infringe too much on what other major moral theories deem centrally important.

#### Existential risk precludes the instantiation of any other ethical theory since individuals cannot reflect upon it, which is a logical prior. In the face of moral uncertainty, the most important thing is preserving our ability to continue deliberating.

## NR

### NR—Yudkowsky

#### Revisionary intuitionism comes before meta-level moral questions about the nature of moral truth—all moral truths fundamentally must be derived from intuition, so if I win that those are flawed you default to util.

#### Specifically, scope insensitivity studies show that people will pay the same amount to save 5,000 lives as 50,000 which means that we should revise our intuition to include aggregation—this means util.

### NR—SideConstraints

#### Side-constraints collapse to consequentialism—if I push someone off a cliff, I can’t know that that violates a side-constraint unless I predict that they die—this means that it is impossible to have a consistent theory of rights without evaluating end states. Thus, we are obligated to positively influence those end states.

### NR—Empirics

#### It is impossible to prescribe action without a consideration of empirical evaluations of how much risk of a rule violation is sufficient to justify moral criticism. Since rational beings follow the course of action that offers the highest probability of producing a desired end, morality must be ends-based to serve its stated purpose.

### NR—Persson

#### The act-omission distinction is nonsensical—if I have initiated an action that will cause death, it would be both permissible and impermissible to let myself kill others.

### NR—Sunstein

#### Morality is a question of what an actor must do in a given situation—to be a just state is to maximize life and minimize violations of side-constraints. The unique nature of a state renders the act-omission distinction incoherent because the state always and necessarily faces a choice between two actions.

### NR—Epistemic Uncertainty

#### You should adopt the parliamentary model for evaluating the framework debate—it’s the only way to resolve moral uncertainty and moral absolutism. Rather than simply affording to 100% risk to one moral theory or another, you should allow theories high leverage on issues important to them—that’s Bostrom. That means that even if I only win 1% risk util is true you should default to preventing existential risk.

### NR—Enoch

### NR—AT: calculation impossible

#### 1. The conclusion does not follow from the premise. These arguments just prove util calc is hard, but that doesn’t mean util is false.

#### 2. A risk of utility is reason enough to take an action, even if we aren’t 100% certain of the consequences.

#### 3. This argument is nonsensical, even if one can’t find the best choice one can be sure that continual calculating is not the best choice, that means unending calculation is inconsistent with utilitarianism.

### NR—AT: infinite timeframe

#### 1. Infinite timeframe doesn’t paralyze util. Even if infinite future possibilities, the probability is infinitely small because they are infinitely far away and there are an infinite number of them.

#### 2. TURN – Infinite timeframe paralyzes <<x theory>>. Under <<x fwk>>, timeframe is irrelevant, so given time there will be a guaranteed violation of sideconstraints on either side.

# F/W—Util—Backend

## 1NC

### NC Util—Yudkowsky

#### The standard is maximizing happiness.

#### Revisionary intuitionism is true and justifies util

Yudkowsky 8 Eliezer Yudkowsky (research fellow of the Machine Intelligence Research Institute; he also writes Harry Potter fan fiction). “The ‘Intuitions’ Behind ‘Utilitarianism.’” 28 January 2008. LessWrong. http://lesswrong.com/lw/n9/the\_intuitions\_behind\_utilitarianism/

I haven't said much about metaethics - the nature of morality - because that has a forward dependency on a discussion of the Mind Projection Fallacy that I haven't gotten to yet. I used to be very confused about metaethics. After my confusion finally cleared up, I did a postmortem on my previous thoughts. I found that my object-level moral reasoning had been valuable and my **meta-level moral reasoning had been worse than useless.** And this appears to be a general syndrome - **people do** not do **much better when discussing whether torture is** good or **bad than when they discuss the meaning of "good" and "bad". Thus, I deem it prudent to keep moral discussions on the object level** wherever I possibly can. Occasionally **people object** to any discussion of morality on the grounds **that morality doesn't exist**, and in lieu of jumping over the forward dependency to explain that **"exist" is not the right term to use** here, I generally say, "But **what do you do anyway?**" and **take the discussion back down to the object level.** Paul Gowder, though, has pointed out that both the idea of choosing a googolplex dust specks in a googolplex eyes over 50 years of torture for one person, and the idea of "utilitarianism", depend on "intuition". He says I've argued that the two are not compatible, but charges me with failing to argue for the utilitarian intuitions that I appeal to. Now "intuition" is not how I would describe the computations that underlie human morality and distinguish us, as moralists, from an ideal philosopher of perfect emptiness and/or a rock. But I am okay with using the word "intuition" as a term of art, bearing in mind that "intuition" in this sense is not to be contrasted to reason, but is, rather, the cognitive building block out of which both long verbal arguments and fast perceptual arguments are constructed. **I see** the project of **morality as a project of renormalizing intuition.** We have intuitions about things that seem desirable or undesirable, intuitions about actions that are right or wrong, intuitions about how to resolve conflicting intuitions, intuitions about how to systematize specific intuitions into general principles. **Delete all** the **intuitions, and** you aren't left with an ideal philosopher of perfect emptiness, **you're left with a rock. Keep all your** specific **intuitions and** refuse to build upon the reflective ones, and you aren't left with an ideal philosopher of perfect spontaneity and genuineness, **you're left with a** grunting **caveperson** running in circles, due to cyclical preferences and similar inconsistencies. "Intuition", as a term of art, is not a curse word when it comes to morality - there is nothing else to argue from. **Even modus ponens is an "intuition"** in this sense - **it**'s **just** that modus ponens **still seems like a good idea after being** formalized, **reflected on**, extrapolated out to see if it has sensible consequences, etcetera. So that is "intuition". However, Gowder did not say what he meant by "utilitarianism". Does utilitarianism say... That right actions are strictly determined by good consequences? That praiseworthy actions depend on justifiable expectations of good consequences? That probabilities of consequences should normatively be discounted by their probability, so that a 50% probability of something bad should weigh exactly half as much in our tradeoffs? That virtuous actions always correspond to maximizing expected utility under some utility function? That two harmful events are worse than one? That two independent occurrences of a harm (not to the same person, not interacting with each other) are exactly twice as bad as one? That for any two harms A and B, with A much worse than B, there exists some tiny probability such that gambling on this probability of A is preferable to a certainty of B? If you say that I advocate something, or that my argument depends on something, and that it is wrong, do please specify what this thingy is... anyway, I accept 3, 5, 6, and 7, but not 4; I am not sure about the phrasing of 1; and 2 is true, I guess, but phrased in a rather solipsistic and selfish fashion: you should not worry about being praiseworthy. Now, what are the "intuitions" upon which my "utilitarianism" depends? This is a deepish sort of topic, but I'll take a quick stab at it. First of all, it's not just that someone presented me with a list of statements like those above, and I decided which ones sounded "intuitive". Among other things, **if you try to violate** "**util**itarianism", **you run into paradoxes, contradictions**, circular preferences, **and other** things that aren't **symptoms of** moral wrongness so much as **moral incoherence.** After you think about moral problems for a while, and also find new truths about the world, and even discover disturbing facts about how you yourself work, you often end up with different moral opinions than when you started out. This does not quite define moral progress, but it is how we experience moral progress. As part of my experienced moral progress, I've drawn a conceptual separation between questions of type Where should we go? and questions of type How should we get there? (Could that be what Gowder means by saying I'm "utilitarian"?) The question of where a road goes - where it leads - you can answer by traveling the road and finding out. If you have a false belief about where the road leads, this falsity can be destroyed by the truth in a very direct and straightforward manner. When it comes to wanting to go to a particular place, this want is not entirely immune from the destructive powers of truth. You could go there and find that you regret it afterward (which does not define moral error, but is how we experience moral error). But, even so, wanting to be in a particular place seems worth distinguishing from wanting to take a particular road to a particular place. Our intuitions about where to go are arguable enough, but our intuitions about how to get there are frankly messed up. **After** the two hundred and eighty-seventh **research** study **showing that people will chop their own feet off if you frame the problem the wrong way, you start to distrust first impressions. When you've read enough research on scope insensitivity** - people will pay only 28% more to protect all 57 wilderness areas in Ontario than one area, **people will pay the same amount to save 50,000 lives as 5,000 lives**... that sort of thing... Well, the worst case of scope insensitivity I've ever heard of was described here by Slovic: Other recent research shows similar results. Two Israeli psychologists asked people to contribute to a costly life-saving treatment. They could offer that contribution to a group of eight sick children, or to an individual child selected from the group. The target amount needed to save the child (or children) was the same in both cases. Contributions to individual group members far outweighed the contributions to the entire group. There's other research along similar lines, but I'm just presenting one example, 'cause, y'know, eight examples would probably have less impact. If you know the general experimental paradigm, then the reason for the above behavior is pretty obvious - focusing your attention on a single child creates more emotional arousal than trying to distribute attention around eight children simultaneously. So people are willing to pay more to help one child than to help eight. Now, **you could** look at this intuition, and **think it was** revealing **some** kind of **incredibly deep moral truth** which shows that one child's good fortune is somehow devalued by the other children's good fortune. But what about the billions of other children in the world? Why isn't it a bad idea to help this one child, when that causes the value of all the other children to go down? How can it be significantly better to have 1,329,342,410 happy children than 1,329,342,409, but then somewhat worse to have seven more at 1,329,342,417? **Or you could** look at that and **say: "The intuition is wrong: the brain can't** successfully **multiply** by eight and get a larger quantity than it started with. **But it ought to**, normatively speaking." And once you realize that the brain can't multiply by eight, then the other cases of scope neglect stop seeming to reveal some fundamental truth about 50,000 lives being worth just the same effort as 5,000 lives, or whatever. You don't get the impression you're looking at the revelation of a deep moral truth about nonagglomerative utilities. It's just that the brain doesn't goddamn multiply. Quantities get thrown out the window. If you have $100 to spend, and you spend $20 each on each of 5 efforts to save 5,000 lives, you will do worse than if you spend $100 on a single effort to save 50,000 lives. Likewise if such choices are made by 10 different people, rather than the same person. As soon as you start believing that it is better to save 50,000 lives than 25,000 lives, that simple preference of final destinations has implications for the choice of paths, when you consider five different events that save 5,000 lives. (It is a general principle that Bayesians see no difference between the long-run answer and the short-run answer; you never get two different answers from computing the same question two different ways. But the long run is a helpful intuition pump, so I am talking about it anyway.) The aggregative valuation strategy of "shut up and multiply" arises from the simple preference to have more of something - to save as many lives as possible - when you have to describe general principles for choosing more than once, acting more than once, planning at more than one time. Aggregation also arises from claiming that the local choice to save one life doesn't depend on how many lives already exist, far away on the other side of the planet, or far away on the other side of the universe. Three lives are one and one and one. No matter how many billions are doing better, or doing worse. 3 = 1 + 1 + 1, no matter what other quantities you add to both sides of the equation. And if you add another life you get 4 = 1 + 1 + 1 + 1. That's aggregation. **When you've read enough** heuristics and **biases research, and enough coherence** and uniqueness **proofs for** Bayesian probabilities and **expected utility**, and you've seen the "Dutch book" and "money pump" effects that penalize trying to handle uncertain outcomes any other way, **then you don't see** the **preference reversals** in the Allais Paradox **as** revealing **some** incredibly **deep moral truth** about the intrinsic value of certainty. **It just goes to show that the brain doesn't** goddamn **multiply.** The primitive, perceptual intuitions that make a choice "feel good" don't handle probabilistic pathways through time very skillfully, especially when the probabilities have been expressed symbolically rather than experienced as a frequency. So you reflect, devise more trustworthy logics, and think it through in words. When you see people insisting that no amount of money whatsoever is worth a single human life, and then driving an extra mile to save $10; or when you see people insisting that no amount of money is worth a decrement of health, and then choosing the cheapest health insurance available; then you don't think that their protestations reveal some deep truth about incommensurable utilities. Part of it, clearly, is that **primitive intuitions don't successfully diminish the emotional impact of** symbols standing for **small quantities** - anything you talk about seems like "an amount worth considering". And part of it has to do with preferring unconditional social rules to conditional social rules. Conditional rules seem weaker, seem more subject to manipulation. If there's any loophole that lets the government legally commit torture, then the government will drive a truck through that loophole. So it seems like there should be an unconditional social injunction against preferring money to life, and no "but" following it. Not even "but a thousand dollars isn't worth a 0.0000000001% probability of saving a life". Though the latter choice, of course, is revealed every time we sneeze without calling a doctor. The rhetoric of sacredness gets bonus points for seeming to express an unlimited commitment, an unconditional refusal that signals trustworthiness and refusal to compromise. So you conclude that moral rhetoric espouses qualitative distinctions, because espousing a quantitative tradeoff would sound like you were plotting to defect. On such occasions, people vigorously want to throw quantities out the window, and they get upset if you try to bring quantities back in, because quantities sound like conditions that would weaken the rule. But you don't conclude that there are actually two tiers of utility with lexical ordering. You don't conclude that there is actually an infinitely sharp moral gradient, some atom that moves a Planck distance (in our continuous physical universe) and sends a utility from 0 to infinity. You don't conclude that utilities must be expressed using hyper-real numbers. Because the lower tier would simply vanish in any equation. It would never be worth the tiniest effort to recalculate for it. All decisions would be determined by the upper tier, and all thought spent thinking about the upper tier only, if the upper tier genuinely had lexical priority. As Peter Norvig once pointed out, if Asimov's robots had strict priority for the First Law of Robotics ("A robot shall not harm a human being, nor through inaction allow a human being to come to harm") then no robot's behavior would ever show any sign of the other two Laws; there would always be some tiny First Law factor that would be sufficient to determine the decision. Whatever value is worth thinking about at all, must be worth trading off against all other values worth thinking about, because thought itself is a limited resource that must be traded off. When you reveal a value, you reveal a utility. I don't say that morality should always be simple. I've already said that the meaning of music is more than happiness alone, more than just a pleasure center lighting up. I would rather see music composed by people than by nonsentient machine learning algorithms, so that someone should have the joy of composition; I care about the journey, as well as the destination. And I am ready to hear if you tell me that the value of music is deeper, and involves more complications, than I realize - that the valuation of this one event is more complex than I know. But that's for one event. When it comes to multiplying by quantities and probabilities, complication is to be avoided - at least if you care more about the destination than the journey. **When you've reflected** on enough intuitions, **and corrected enough absurdities, you** start to **see a common denominator, a meta-principle** at work, **which one might phrase as "Shut up and multiply."** Where music is concerned, I care about the journey. When lives are at stake, I shut up and multiply. It is more important that lives be saved, than that we conform to any particular ritual in saving them. And the optimal path to that destination is governed by laws that are simple, because they are math. **And that's why I'm a utilitarian** - at least when I am doing something that is overwhelmingly more important than my own feelings about it - which is most of the time, because there are not many utilitarians, and many things left undone.

### NC Util—Parfit

Brain studies prove personal identity doesn’t exist. **Parfit 84** writes[[2]](#footnote-2)

Some **recent medical cases provide striking evidence in favour of the Reductionist View.** Human beings have a **lower brain and** two **upper hemispheres**, which **are connected by a bundle of fibres.** In treating a few people with severe epilepsy, **surgeons have cut these fibres.** The aim was to reduce the severity of epileptic fits, by confining their causes to a single hemisphere. This aim was achieved. But the operations had another unintended consequence. **The effect**, in the words of one surgeon, **was the creation of ‘two separate spheres of consciousness.’ This effect was revealed by** various **psychological tests.** These made use of two facts. We control our right arms with our left hemispheres, and vice versa. And what is in the right halves of our visual fields we see with our left hemispheres, and vice versa. When someone’s hemispheres have been disconnected, **psychologists can thus present** to this person two different written **questions in the two halves of his visual field, and can receive two different answers** written by this person’s two hands.

In the absence of personal identity, only end states can matter. **Shoemaker 99**[[3]](#footnote-3)

Extreme reductionism might lend support to utilitarianism in the following way. Many people claim that we are justified in maximizing the good in our own lives, but not justified in maximizing the good across sets of lives, simply because each of us is a single, deeply unified person, unified by the further fact of identity, whereas there is no such corresponding unity across sets of lives. But if the only justification for the different treatment of individual lives and sets of lives is the further fact, and this fact is undermined by the truth of reductionism, then nothing justifies this different treatment. **There are no deeply unified subjects of experience. What remains are merely the experiences themselves, and so any ethical theory distinguishing between individual lives** and sets of lives **is mistaken.** If the deep, further fact is missing, then there are no unities. **The morally significant units should then be the states people are in at particular times, and an ethical theory that focused on them** and attempted to improve their quality, whatever their location, **would be the most plausible. Util**itarianism **is just such a theory.**

### NC Util—Rationality

#### Consequentialist theories are the simplest explanation of rational decision-making

Pettit 99 Philip Pettit (Laurence S. Rockefeller University Professor of Politics and Human Values at Princeton). “Consequentialism.” A Companion to Ethics, ed. Peter Singer. Oxford. 1991.

There are at least three respects in which **consequentialism scores on simplicity**. The first is that whereas consequentialists endorse only one way of responding to values, non-consequentialists endorse two. **Non-consequentialists** all **commit** themselves **to the view that certain values should be honoured rather than promoted**: say, values like those associated with loyalty and respect. **But they** all **agree**, whether or not in their role as moral theorists, that **certain** other **values should be promoted**: values **as various as economic prosperity, personal hygiene, and** the **safety of nuclear installations**. Thus **where consequentialists introduce a single axiom on how values justify choices, non-consequentialists must introduce two.** But not only is non-consequentialism less simple for losing the numbers game. It is also less simple for playing the game in an ad hoc way. Non-consequentialists all identify certain values as suitable for honouring rather than promoting. But they do not generally explain what it is about the values identified which means that justification comes from their being honoured rather than promoted. And indeed it is not clear what satisfactory explanation can be provided. **It is one thing to make a list of** the **values which allegedly require honouring**: values, say, like personal loyalty, respect for others, and punishment for wrongdoing. **It is another to say why these values are so** very **different from** the **ordinary** run of **desirable properties**. There may be features that mark them off from other values, but why do those features matter so much? That question typically goes unconsidered by non-consequentialists. Not only do they have a duality then where consequentialists have a unity; **they** also **have an unexplained duality**. The third respect in which **consequentialism** scores on the simplicity count is that it **fits nicely with** our **standard views of** what **rationality** requires, whereas non-consequentialism is in tension with such views. The agent concerned with a value is in a parallel position to that of an agent concerned with some personal good: say, health or income or status. In thinking about how an agent should act on the concern for a personal good, **we unhesitatingly say that** of course **the rational thing to do**, the rationally justified action, **is to act so that the good is promoted**. That means then that whereas the consequentialist line on how values justify choices is continuous with the standard line on rationality in the pursuit of personal goods, the non-consequentialist line is not. **The non-consequentialist has the embarrassment of having to defend a position** on what certain values require which is **without analogue in** the **non-moral** area of **practical rationality.**

### NC Util—AT: ripstein

**Freedom alone can’t justify even basic state health policies which proves that her framework fails. Pain also harms freedom, so Ripstein devolves to util.**

**Yankah 12**

Ekow Yankah (Professor of Law @ Cardozo School of Law; New York democratic politician; holds degrees from UMich, Columbia, and Oxford). “Crime, Freedom, and Civic Bonds: Arthur Ripstein’s Force and Freedom: Kant’s Legal and Political Philosophy.” Criminal Law and Philosophy. Springer. March 4th, 2012.

Because Ripstein’s Kant takes only the protection of freedom to justify law, law is indifferent to the potential harms that befall one outside of the protection of freedom (48). Indeed, Ripstein repeatedly points out that a Kantian legal structure cannot premise legal coercion on any other benefits it can secure (1 85-190). The law, in such a world, is utterly insensitive to any other interests a person might have insofar as they are not freedom preserving (Tadros 2011). This, Ripstein insists, includes even your most fundamental interests, yes, life itself (274-275). Yet Ripstein, like most of us, must be well aware that many, perhaps most of the activities of the thick modern state reach[es] deeply into the lives of citizens in ways that seem aimed at preserving their health, wealth and standards of living. One finds it hard to imagine that Ripstein, the proud Torontonian, would be willing to give up all the public parks in that city to say nothing of the Canadian national healthcare system. The tug of this dilemma will be equally felt by all who are tempted to take freedom not simply as a unique but as the sole justification for a legal system (Tadros 201 l:200-2ol, 205-206). In order to placate, if not entirely satisfy this intuition, Ripstein must pack an awful lot into his concept of freedom. To do so, Ripstein relies on Kant’s notion, earlier raised, that equal freedom can only be secured in a “rightful condition.” Thus, the only powers the state can justifiably wield are those aimed at perfecting the rightful condition (23). Ripstein argues, however, that securing the rightful condition confers broad powers. Some of these state powers are quite naturally described as preserving freedom. As earlier noted, Ripstein argues that freedom allows the state to protect citizens against extreme poverty or reliance only on private charities, for this would make such citizens dependent on (and dominated by) the private will of other individuals. In the same vein, roads are required not for convenience or utility but to allow public movement and thus freedom. Likewise, the modern state may provide for generally for public health in order to secure freedom in the rightful condition by, for example, quarantining and treating potentially dangerous epidemics (240). This strikes me as true as far as it goes but one cannot help to eventually grow suspicious that Ripstein’s concept of freedom has ironically imperial ambitions, seeking to include much of what we have come to think of as perfectly natural functions of the modern state. Let’s take a recent controversial example in America. I think that the connection between healthcare and freedom draws on more than dangerous epidemics which threaten the very integrity of the state. One ought not underestimate the ways in which the recurring illness and chronic pains that beset so many erodes, pinprick by increasingly hobbling pinprick, the ability to engage even basic freedoms. Still, it seems overly ambitious to describe the entire range of modern health care services provided by most developed countries as plausibly justified as merely freedom securing. Modern health care services in Canada, to take one example, quickly outrace the basic services that preserve the basic freedom in order to contain costs and distribute a basic good for the welfare of its citizens. Secondly, it remains unclear to me, despite Ripstein’s confidence, why freedom as such would permit, much less could require, the State to provide health services. This is true for two reasons which undermine his earlier arguments. Perhaps one can understand why having citizens only rely on private charity alone could pose a risk to freedom. All too often, particular charities capture the public imagination, receiving disproportionate amounts of funding, while other problems insufficiently championed by glamorous advocates or afflicting poor, minority or other populations go largely unaided. What is harder to describe as required by freedom alone is a state needing to provide healthcare if there were a functioning private market, even if government intervention could assist others in securing better healthcare outcomes at better prices for all. If this worry is correct and the concept of freedom is being overworked here Ripstein cannot, despite his seeming equanimity on the subject, justify even the simplest government actions that improve health beyond quarantines and crippling disease, to say nothing of public parks, international aid to the most desperate of countries and support for the arts. One worries that Ripstein, married to the idea of freedom as a singular justification for law, reaches to find a way to include aspects of public action that stretch well beyond securing freedom, lest he present too bare a picture of government for most to stomach. Unless Ripstein is willing to sacrifice, indeed prohibit, much of the modern state, in which case the reader should be clearly informed, Kantian freedom begins to look too thin.

Ripstein’s conception of freedom is circular and unintuitive

**Valentini 12** writes[[4]](#footnote-4)

In this section, I argue that **Ripstein’s ‘right to freedom’** cannot ground all other rights because the notion of freedom on which it relies **presupposes** the very **rights it aims to establish**. This is what I call the ‘circle of freedom’. This vicious circularity arises from Ripstein’s endorsement of the following claims: a. The right to freedom grounds all other rights. b. The right to freedom is the right of each individual to be his/her own master, to be independent of the will of others. c. Independence of the will of others consists in the ability to use one’s own means to pursue one’s own purposes robustly unhindered by others. d. One’s own means and purposes are the means and purposes one has a right to. e. The right to freedom is therefore the right to use the means and pursue the purposes one has a right to, robustly unhindered by others. As Ripstein puts it, a system where all have freedom as independence ‘is one in which each person is free to use his or her powers, individually or cooperatively, to set his or her own purposes, and no one is allowed to compel others to use their powers in a way designed to advance or accommodate any other person’s purposes’ (ibid.: 33, emphases added). But how are we to determine what one’s powers and purposes are? Certainly not by looking at their actual powers and purposes. To be sure, **when policemen stop a thief**, they prevent him from using his (positive, as opposed to normative) powers for his (positive) purposes, yet **we would hardly regard such an intervention as** unjust, as a **violation of the thief’s right to freedom**. This is paradigmatically a legitimate intervention, aimed at ‘hindering a hindrance to freedom’ (i.e., the freedom of the victim, whose means would serve someone else’s, the thief’s, purposes). The **freedom referred to in** the expression **‘hindering a hindrance** to freedom**’ cannot be any freedom, but** must be the **freedom one is entitled to on grounds of justice. Until we have an independent account of justice**, then, **we cannot know whether someone is free** or unfree. Unless we know what is ours, we cannot know whether constraints on our de facto agency are violations of our independence or consistent with it. Rather than grounding all rights and entitlements, Ripstein’s Kantian notion of freedom is derivative of them (i.e., it presupposes them). This appears clear once we notice that the cases Ripstein offers to illustrate instances of dependence and independence only work for his purposes if we assume a certain background account of justice. For instance, in the example offered earlier, involving market competition between Sam and John, a tacit assumption was made about the entitlement-generating character of free market processes. Recall that, in Ripstein’s view, Sam’s driving customers away from John does not constitute a violation of John’s freedom as independence. This can only be so on the assumption that free market exchanges are entitlementgenerating independently of their outcomes. This assumption is controversial, and certainly not ‘implicit’ in the meaning of freedom. On some accounts of justice (Rawls’s, for instance), free market processes need to be regulated in order to be consistent with individuals’ rights. If such processes lead to excessive inequalities, Rawls argues, their outcomes need to be rectified in order to preserve free market exchanges over time (Rawls 1993: 266).6 Whether the interaction between Sam and John involves a breach of freedom as independence, then, depends on what particular account of rights and entitlements one holds. The right to freedom as independence is not the answer, but an independent (and necessarily controversial) account of persons’ rights is needed to know what freedom as independence is. If my argument up to this point is correct, the unified nature of the Kantian approach offered by Ripstein is only illusory. His articulation of the right to freedom cannot constitute the ground of all other rights because freedom itself is defined in terms of persons’ rights. Without a prior account of what those rights are, the notion of freedom as independence is empty; with such an account, it is expositionally parsimonious, but surreptitiously presupposes a complex theory of justice. I have suggested that Ripstein’s articulation of the notion of freedom presupposes an account of individual rights and thus cannot strictly speaking ground any such rights. Despite its lacking rights-grounding capacity, this notion may still be of value. That is, it may offer a plausible account of freedom, which we might want to employ in elaborating our all-things-considered theory of persons’ rights and entitlements. After all, as we saw earlier, this notion is more in line with at least some of our intuitive judgements about freedom than the popular notion of freedom as non-interference.7 Freedom as independence conceives of persons’ freedom in relation to their in-principle subjection (or lack thereof) to the will of others. Recall that a slave with a benevolent master is still unfree because in principle subject to the master’s will. Even though the master does not interfere with the slave in the actual world, there are many nearby possible worlds in which such interference would occur (the master is indeed legally entitled to interfere with the slave), and this fact, says the proponent of freedom as independence, must be taken into account when judging whether the slave is free (cf. the discussion in Pettit 1997: ch. 2, and List 2006). Although such a focus on the robustness of non-interference renders freedom as independence rather appealing, **the appeal is significantly undermined by** this notion’s **reliance on a prior conception of rights. If to be independent** of the will of another **is to not have one’s rights violated** (robustly across possible worlds), then limitations of one’s capacity to act that do not violate rights do not count as restrictions of freedom. On this view, **my freedom is not restricted when I am not allowed to access property that is not mine**.8 Or else, my freedom is not restricted whenever I am forced to pay taxes (if such taxes are demanded by justice). Even more strikingly, **I cannot say that my freedom is restricted if I am justly incarcerated for violating** others’ **rights**. All of **these judgements are** deeply **counter-intuitive**, but they inevitably follow from an understanding of freedom according to which someone is free if she can robustly use the means and pursue the purposes she has a right to use and pursue. What we would intuitively call ‘justified’ restrictions of freedom are no restrictions of freedom at all, on Ripstein’s account.9 It is worth noting at this point that these counter-intuitive implications of freedom as independence are not fully transparent from Ripstein’s text. In fact, there are passages, discussing the use of coercion, which explicitly exclude them. Ripstein tells us that ‘Kant does not conceive of coercion in terms of threats, but instead as the limitation of freedom’ (Ripstein 2009: 54). From this it would seem to follow that acts of coercion that are consistent with freedom (i.e., with people’s rights) simply do not count as coercive because they do not limit freedom. Again, forcing a criminal to go to jail, on this view, would not be ‘coercive’ because it would be consistent with his freedom as independence (i.e., the freedom he has a right to). Yet Ripstein does not use the language of coercion in this way. Instead, he distinguishes between legitimate and illegitimate coercion, the former being coercion exercised in accordance with people’s rights, the latter being coercion exercised in breach of those rights. He illustrates this with the following example: Using force to get the victim out of the kidnapper’s clutches involves coercion against the kidnapper, because it touches or threatens to touch him in order to advance a purpose, the freeing of the victim, to which Kant, Ripstein, and the Circle of Freedom: A Critical Note 455 © 2012 Blackwell Publishing Ltd. he has not agreed. The use of force is rightful because an incident of the victim’s antecedent right to be free. (ibid.: 55) In this quote, Ripstein appeals to a notion of freedom which differs from the moralized one we encountered in the previous section. If it is true that the use of force to free the victim limits the kidnapper’s freedom because it prevents him from using his resources to achieve his purposes, then ‘his resources’ and ‘his purposes’ have to be interpreted in positive rather than normative terms. ‘His’ resources and purposes are not those he has a right to, but those he happens to possess. There thus appear to be two notions of freedom at play in Ripstein’s work, one (the dominant one, it seems to me) is moralized, the other non-moralized10: FMoralized = A is free if, and only if, A can use the means and pursue the purposes A has a right to, robustly unhindered by others. FNon-Moralized = A is free if, and only if, A can use the means A happens to possess and pursue the purposes A happens to have, robustly unhindered by others.

### NC Util—AT: Contractualism

#### Contractualism fails at providing a source of moral motivation and the source it does provide fails to treat humans as intrinsically valuable

Shoemaker 2k (David Shoemaker (California State University). “Reductionist Contractualism: Moral Motivation and the Expanding Self.” 2000. http://www.csun.edu/~ds56723/cjppaper.pdf)

But is this in fact the most plausible account of all moral motivation? And even if we have reason to think that it is, how might we convince the amoralist to cultivate the motivational desire at the heart of this account? The first problem has been articulated in a number of articles attacking both utilitarian and deontological views (such as Kantian contractualism) for ignoring the sources of moral motivation found in human beings’ pursuit of the good life. 9 Against the kind of contractual- ism I have discussed, the objection might run as follows. Surely it must be agreed that some of the greatest human goods include love, friend- ship, community attachments, and the having of one or another ground project (a project the pursuit of which lends meaning to one’s life). And having these sorts of goods necessarily involves having certain motives for action. For example, the love I have for my wife necessarily involves not only valuing her qua beloved, but also not valuing her as, say, simply the possessor or producer of certain values (e.g., as a particular instan- tiation here and now of a free moral agent) (Stocker, 459). So I am moved to help her, for example, simply in light of the fact that she is my beloved. Similarly, I may have certain ground projects which provide meaning for my life, and it is these ground projects which propel my life forward and give me reasons for acting in certain ways (Williams, 10-15). But it is alleged that **the source of moral motivation elaborated by the contractualist seems** importantly **at odds with** the source of **motivation underlying** these **everyday moral pursuits**, in one of two ways. On the one hand, it has been claimed that if the ultimate source of moral motivation is to be the desire for justifiability, then the very possibility of pursuing certain of these crucial human goods is undermined. After all, **if my motivational set embodies** the notion of **justification as its sole focus, I will** necessarily **be treating all other people externally**, that is, **as possessors of certain values, and not** as **intrinsically valuable,** that is, as valuable in themselves, **a stance required for love, friendship, and** certain **community attachments** (Stocker, 461). **If I am moved** to action **solely by** the **desire for justifiability,** then **this motivation seems to preclude one of the motivations involved in love, e.g., that I perform** certain **actions entirely for the sake of the beloved** and for no other reason (such as, for example, that the beloved is a person for whom the notion of such justification makes sense). On the other hand, suppose contractualists are not claiming that the desire for justifiability is the only source of moral motivation. Perhaps it is merely the ‘trumping’ source. 10 Yet even if this is the case, a similar problem remains. For it seems as if we all recognize ‘that among the immensely valuable traits and activities that a human life might posi- tively embrace are some of which we hope that, if a person does embrace them, he does so not for moral reasons’ (Wolf, 434). Suppose, for example, that I am on a boat with my wife and a stranger. The boat capsizes, and I am able to rescue only one of these two people. What would the contractualist have us do? It would surely be absurd to maintain impar- tiality by flipping a coin, and fortunately the Scanlonian picture avoids such a result. Rather, it would be permissible for me to save my wife just in case no one could reasonably reject a set of rules allowing such an action, and in this case it seems as if no one could reasonably reject such a set of rules. So I can justify to the stranger my letting him die on grounds he himself could reasonably accept. But notice the oddness of the result, for it sounds as if some relief is attached: ‘Whew! Thank goodness my moral commitments didn’t require me to let my wife die!’ 11 The oddness here, as Bernard Williams points out, stems from there being one thought too many. One would think that, finding myself in such circumstances, I would be motivated to save my wife solely because she is my wife , and not because she is my wife and in situations of this kind it is permissible o save one’s wife’’’ (ibid., 18). Some morally charged situations, it seems, lie outside the arena of justification altogether. Consequently, **the contractualist faces a real problem if she cannot fully account for** the **motivational source(s) grounding** these **crucial human pursuits**. On the one hand, if the Scanlonian source of moral motivation is supposed to be our sole ultimate source of moral motiva- tion, then it would have to preclude the possibility of genuine love and fellow-feeling, which is absurd. On the other hand, **even if** this desire for **justifiability is merely** to be **the trumping motivational source** among certain other desires or reasons for action, then **it still occasionally provides one thought too many and thus seems** at worst implausible and **at best redundant**. The desire for justifiability thus looks inadequate as a source of moral motivation, and any contractualism resting on such a desire would seem to be highly problematic.

#### Contractualism devolves to util.

Cummiskey 08 (“Dignity, Contractualism, and Consequentialism,” David Cummiskey. Utilitas Vol. 20, No. 4, December 2008.)

**Scanlon’s contractualism** should set aside at the start agent-relative *moral* reasons, that is, special obligations and deontological reasons, and start only with the non-moral agent-relative reasons of autonomy. On this revised interpretation, the grounds for reasonable rejection can **appeal[s]** only **to the fundamental projects and commitments of the contractors.** The problem for the deontologist, however, remains. We still need an explanation of how this procedure gives rise to agent-relative ‘reasons of special obligation’ and ‘deontological reasons’. Such a claim is utterly *ad hoc*, unmotivated and unjustified. Since the only agent-relative reasons that are inputs to the contractualist procedure are ‘reasons of autonomy’ – individual commitments to projects, relationships and other personal goals – the moral reasons must piggyback on these types of reasons alone. One obvious proposal is that reasonable grounds for the rejection of principles must be based on the assumption that there is a corresponding agent-neutral reason for every agent-relative reason. **The result would be a principle of reasonable rejection that gives impartial equal consideration or weight to each agent’s ends.** So interpreted, it is hard to see why **this** does not **result[s] in** some form of consequentialist principle. What we have here is a familiar argument for **util**itarianism – with the standard concerns over whether the fundamental principle will be a maximizing or satisficing principle, and of course whether it will be simply aggregative or also include distribution-sensitive considerations based on the priority of theworst-off or egalitarian values.24 In short, contractualism so interpreted seems to generate some form of aggregative and/or distribution-sensitive consequentialism. I believe that the closest Scanlon comes to addressing this alternative is in his discussion of ‘welfare contractualism’. His rejection of welfare contractualism, however, brings us back to the problem with which we started.

### NC Util—AT: rawls

#### 1. Util controls the link to Rawls. Under the veil of ignorance, you would agree to maximize utility to increase the chance of one’s own interests being satisfied.

#### 2. Extinction turns Rawls. It harms the least well-off.

#### 3. Rawls devolves to util. Otherwise his assumptions about primary goods fail.

Arneson 00 (Professor in the Department of Philosophy at the University of California, San Diego)

The defender of a Rawlsian primary goods standard might interject a skeptical response at this point: the utility-based conceptions of justice are nonstarters because no satisfactory standard of interpersonal comparison for the purpose can be developed. In a Theory of Justice, Rawls does not emphasize the difficulties about interpersonal comparison, and indicates the superiority of justice as fairness lies along some other dimension. The interpersonal comparison problem is no doubt significant, though in my judgment not insoluble. Here I wish to make a more limited point: **Rawls** has his own unsolved difficulties with interpersonal comparison, so, **living in a glass house,** he **is poorly placed to** be **throw**ing **stones at util**itarian **windows. According to Rawls, there are several primary** social **goods**. Some of these qualify as basic liberties, so are treated within the equal liberty principle, which is accorded strict priority within Rawls’s system. But this leaves several primary goods **other than basic liberties, so** in order **to apply the difference principle**, which requires that the expectations of these primary goods be maximized for the worst off class of persons, **we need an index of these goods**: a way of determining, for any disaparate bundles of these goods, which contain more primary goods overall. If one bundle containing various amounts of various primary goods can be matched with another bundle that contains more of each of these distinct primary goods, then the second bundle dominates the first, and unambiguously contains more primary goods overall. But for the many cases where dominance does not hold, we need an index. **I do not see how** an individual’s bundle of **primary goods can be assessed except in terms of the extent** to which **those goods enable the individual to satisfy her preferences, or** to **fulfill some** given **objective conception of the good**, and neither of these ways of assessment provides a measure that is consistent with Rawls’s strategy of argument and core assumptions. To my knowledge, the only serious discussion of the index problem for primary goods is in John Roemer, Theories of Distributive Justice. 2 Roemer asserts that the index problem admits of a solution, but his proposed solution compromises Rawls’s avoidance of utility-based measures, and causes Rawls’s principles to unravel, so this is not a friendly construal of Rawls that could be used to defend his position against a utilitarian critic. Without pressing this issue further, at the very least there is a problem here that the defender of a primary goods standard would need to solve, and has not solved to date, if the primary goods approach to justice comparison issues is to be a viable position.

### NC Util—Deon -> Util

#### Respect for humans as ends justifies util.

Cummiskey 90 (Cummiskey, David. Associate professor of philosophy at the University of Chicago. “Kantian Consequentiaism.” Ethics 100 (April 1990), University of Chicago. <http://www.jstor.org/stable/2381810>)

We must not obscure the issue by characterizing this type of case as the sacrifice of individuals for some abstract “social entity.” It is not a question of some persons having to bear the cost for some elusive “overall social good.” Instead, the question is whether some persons must bear the inescapable cost for the sake of other persons. Robert Nozick, for example, argues that “to use a person in this way does not sufficiently respect and take account of the fact that he is a separate person, that his is the only life he has.” But why is this not equally true of all those whom we do not save through our failure to act? By emphasizing solely the one who must bear the cost if we act, we fail to sufficiently respect and take account of the many other separate persons, each with only one life, who will bear the cost of our inaction. In such a situation, what would a conscientious Kantian agent, an agent motivated by the unconditional value of rational beings, choose? A morally good agent recognizes that the basis of all particular duties is the principle that “rational nature exists as an end in itself”. Rational nature as such is the supreme objective end of all conduct. If one truly believes that all rational beings have an equal value, then the rational solution to such a dilemma involves maximally promoting the lives and liberties of as many rational beings as possible. In order to avoid this conclusion, the non-consequentialist Kantian needs to justify agent-centered constraints. As we saw in chapter 1, however, even most Kantian deontologists recognize that agent-centered constraints require a non- value-based rationale. But we have seen that Kant’s normative theory is based on an unconditionally valuable end. How can a concern for the value of rational beings lead to a refusal to sacrifice rational beings even when this would prevent other more extensive losses of rational beings? If the moral law is based on the value of rational beings and their ends, then what is the rationale for prohibiting a moral agent from maximally promoting these two tiers of value? If I sacrifice some for the sake of others, I do not use them arbitrarily, and I do not deny the unconditional value of rational beings. Persons may have “dignity, that is, an unconditional and incomparable worth” that transcends any market value, but persons also have a fundamental equality that dictates that some must sometimes give way for the sake of others. The concept of the end-in-itself does not support the view that we may never force another to bear some cost in order to benefit others.

### NC Util—AT: 2nd person morality

#### Second-personal morality devolves to util

Singer 93 (Peter Singer, “Practical Ethics,” Second Edition, Cambridge University Press, 1993, pp. 13-14)

The universal aspect of ethics, I suggest, does provide a persuasive, although not conclusive, reason for taking a broadly utilitarian position. My reason for suggesting this is as follows. **In accepting that ethical judgments must be** made from a **universal** point of view, **I am accepting that my own interests cannot,** simply because they are my interests, **count more than the interests of anyone else. Thus my** very natural **concern that my own interests be looked after must**, when I think ethically, **be extended to** the interests of **others.** Now, imagine that I am trying to decide between two possible courses of action – perhaps whether to eat all the fruits I have collected myself, or to share them with others. Imagine, too, that I am deciding in a complete ethical vacuum, that I know nothing of any ethical considerations – I am, we might say, in a pre-ethical stage of thinking. How would I make up my mind? One thing that would be still relevant would be how the possible courses of action will affect my interests. Indeed, if we define ‘interests’ broadly enough, so that we count anything people desire as in their interests (unless it is incompatible with another desire or desires), then it would seem that at this pre-ethical stage, only one’s own interests can be relevant to the decision. Suppose I then begin to think ethically, to the extent of recognizing that my own interests cannot count for more, simply because they are my own, than the interests of others. In place of my own interests, I now have to take into account the interests of all those affected by my decision. **This requires me to weigh** up **all** these **interests and** adopt the course of action most likely to **maximize the interests of those affected.**

## NR

# Cards

### Goodin—policymaker -> util

#### Util is the only moral system available to policy-makers.

Goodin 90 (Robert Goodin, fellow in philosophy, Australian National Defense University, THE UTILITARIAN RESPONSE, 1990, p. 141-2)

My larger argument turns on the proposition that there is something special about the situation of public officials that makes utilitarianism more probable for them than private individuals. Before proceeding with the large argument, I must therefore say what it is that makes it so special about public officials and their situations that make it both more necessary and more desirable for them to adopt a more credible form of utilitarianism. Consider, first, the argument from necessity. **Public officials** are obliged to **make** their **choices under uncertainty**, and uncertainty **of a** very **special sort** at that. All choices – public and private alike – are made under some degree of uncertainty, of course. But in the nature of things, private individuals will usually have more complete information on the peculiarities of their own circumstances and on the ramifications that alternative possible choices might have for them. Public officials, in contrast, **[they] are** relatively **poorly informed as to the effects that their choices will have on individuals, one by one. What they** typically **do know are** generalities: **averages and aggregates. They know what will happen most often to most people** as a result of their various possible choices, **but that is all. That** is enough to **allow[s]** public **policy-makers to use** the **util**itarian **calculus** – assuming they want to use it at all – to chose general rules or conduct.

#### Prefer state-specific justifications because of actor-specificity—most contextual to the resolutional actor.

### Yudkowsky—Revisionary Intutionism

#### Revisionary intuitionism is true.

Yudkowsky 8 Eliezer Yudkowsky (research fellow of the Machine Intelligence Research Institute; he also writes Harry Potter fan fiction). “The ‘Intuitions’ Behind ‘Utilitarianism.’” 28 January 2008. LessWrong. http://lesswrong.com/lw/n9/the\_intuitions\_behind\_utilitarianism/

I haven't said much about metaethics - the nature of morality - because that has a forward dependency on a discussion of the Mind Projection Fallacy that I haven't gotten to yet. I used to be very confused about metaethics. After my confusion finally cleared up, I did a postmortem on my previous thoughts. I found that my object-level moral reasoning had been valuable and my **meta-level moral reasoning had been worse than useless.** And this appears to be a general syndrome - **people do much better when discussing whether torture is** good or **bad than when they discuss the meaning of "good" and "bad". Thus, I deem it prudent to keep moral discussions on the object level** wherever I possibly can. Occasionally **people object** to any discussion of morality on the grounds **that morality doesn't exist**, and in lieu of jumping over the forward dependency to explain that **"exist" is not the right term to use** here, I generally say, "But **what do you do anyway?**" and **take the discussion back down to the object level.** Paul Gowder, though, has pointed out that both the idea of choosing a googolplex dust specks in a googolplex eyes over 50 years of torture for one person, and the idea of "utilitarianism", depend on "intuition". He says I've argued that the two are not compatible, but charges me with failing to argue for the utilitarian intuitions that I appeal to. Now "intuition" is not how I would describe the computations that underlie human morality and distinguish us, as moralists, from an ideal philosopher of perfect emptiness and/or a rock. But I am okay with using the word "intuition" as a term of art, bearing in mind that "intuition" in this sense is not to be contrasted to reason, but is, rather, the cognitive building block out of which both long verbal arguments and fast perceptual arguments are constructed. **I see** the project of **morality as a project of renormalizing intuition.** We have intuitions about things that seem desirable or undesirable, intuitions about actions that are right or wrong, intuitions about how to resolve conflicting intuitions, intuitions about how to systematize specific intuitions into general principles. **Delete all** the **intuitions, and** you aren't left with an ideal philosopher of perfect emptiness, **you're left with a rock. Keep all your** specific **intuitions and** refuse to build upon the reflective ones, and you aren't left with an ideal philosopher of perfect spontaneity and genuineness, **you're left with a** grunting **caveperson** running in circles, due to cyclical preferences and similar inconsistencies. "Intuition", as a term of art, is not a curse word when it comes to morality - there is nothing else to argue from. **Even modus ponens is an "intuition"** in this sense - **it**'s **just** that modus ponens **still seems like a good idea after being** formalized, **reflected on**, extrapolated out to see if it has sensible consequences, etcetera. So that is "intuition". However, Gowder did not say what he meant by "utilitarianism". Does utilitarianism say... That right actions are strictly determined by good consequences? That praiseworthy actions depend on justifiable expectations of good consequences? That probabilities of consequences should normatively be discounted by their probability, so that a 50% probability of something bad should weigh exactly half as much in our tradeoffs? That virtuous actions always correspond to maximizing expected utility under some utility function? That two harmful events are worse than one? That two independent occurrences of a harm (not to the same person, not interacting with each other) are exactly twice as bad as one? That for any two harms A and B, with A much worse than B, there exists some tiny probability such that gambling on this probability of A is preferable to a certainty of B? If you say that I advocate something, or that my argument depends on something, and that it is wrong, do please specify what this thingy is... anyway, I accept 3, 5, 6, and 7, but not 4; I am not sure about the phrasing of 1; and 2 is true, I guess, but phrased in a rather solipsistic and selfish fashion: you should not worry about being praiseworthy. Now, what are the "intuitions" upon which my "utilitarianism" depends? This is a deepish sort of topic, but I'll take a quick stab at it. First of all, it's not just that someone presented me with a list of statements like those above, and I decided which ones sounded "intuitive". Among other things, **if you try to violate** "**util**itarianism", **you run into paradoxes, contradictions**, circular preferences, **and other** things that aren't **symptoms of** moral wrongness so much as **moral incoherence.** After you think about moral problems for a while, and also find new truths about the world, and even discover disturbing facts about how you yourself work, you often end up with different moral opinions than when you started out. This does not quite define moral progress, but it is how we experience moral progress. As part of my experienced moral progress, I've drawn a conceptual separation between questions of type Where should we go? and questions of type How should we get there? (Could that be what Gowder means by saying I'm "utilitarian"?) The question of where a road goes - where it leads - you can answer by traveling the road and finding out. If you have a false belief about where the road leads, this falsity can be destroyed by the truth in a very direct and straightforward manner. When it comes to wanting to go to a particular place, this want is not entirely immune from the destructive powers of truth. You could go there and find that you regret it afterward (which does not define moral error, but is how we experience moral error). But, even so, wanting to be in a particular place seems worth distinguishing from wanting to take a particular road to a particular place. Our intuitions about where to go are arguable enough, but our intuitions about how to get there are frankly messed up. **After** the two hundred and eighty-seventh **research** study **showing that people will chop their own feet off if you frame the problem the wrong way, you start to distrust first impressions. When you've read enough research on scope insensitivity** - people will pay only 28% more to protect all 57 wilderness areas in Ontario than one area, **people will pay the same amount to save 50,000 lives as 5,000 lives**... that sort of thing... Well, the worst case of scope insensitivity I've ever heard of was described here by Slovic: Other recent research shows similar results. Two Israeli psychologists asked people to contribute to a costly life-saving treatment. They could offer that contribution to a group of eight sick children, or to an individual child selected from the group. The target amount needed to save the child (or children) was the same in both cases. Contributions to individual group members far outweighed the contributions to the entire group. There's other research along similar lines, but I'm just presenting one example, 'cause, y'know, eight examples would probably have less impact. If you know the general experimental paradigm, then the reason for the above behavior is pretty obvious - focusing your attention on a single child creates more emotional arousal than trying to distribute attention around eight children simultaneously. So people are willing to pay more to help one child than to help eight. Now, **you could** look at this intuition, and **think it was** revealing **some** kind of **incredibly deep moral truth** which shows that one child's good fortune is somehow devalued by the other children's good fortune. But what about the billions of other children in the world? Why isn't it a bad idea to help this one child, when that causes the value of all the other children to go down? How can it be significantly better to have 1,329,342,410 happy children than 1,329,342,409, but then somewhat worse to have seven more at 1,329,342,417? **Or you could** look at that and **say: "The intuition is wrong: the brain can't** successfully **multiply** by eight and get a larger quantity than it started with. **But it ought to**, normatively speaking." And once you realize that the brain can't multiply by eight, then the other cases of scope neglect stop seeming to reveal some fundamental truth about 50,000 lives being worth just the same effort as 5,000 lives, or whatever. You don't get the impression you're looking at the revelation of a deep moral truth about nonagglomerative utilities. It's just that the brain doesn't goddamn multiply. Quantities get thrown out the window. If you have $100 to spend, and you spend $20 each on each of 5 efforts to save 5,000 lives, you will do worse than if you spend $100 on a single effort to save 50,000 lives. Likewise if such choices are made by 10 different people, rather than the same person. As soon as you start believing that it is better to save 50,000 lives than 25,000 lives, that simple preference of final destinations has implications for the choice of paths, when you consider five different events that save 5,000 lives. (It is a general principle that Bayesians see no difference between the long-run answer and the short-run answer; you never get two different answers from computing the same question two different ways. But the long run is a helpful intuition pump, so I am talking about it anyway.) The aggregative valuation strategy of "shut up and multiply" arises from the simple preference to have more of something - to save as many lives as possible - when you have to describe general principles for choosing more than once, acting more than once, planning at more than one time. Aggregation also arises from claiming that the local choice to save one life doesn't depend on how many lives already exist, far away on the other side of the planet, or far away on the other side of the universe. Three lives are one and one and one. No matter how many billions are doing better, or doing worse. 3 = 1 + 1 + 1, no matter what other quantities you add to both sides of the equation. And if you add another life you get 4 = 1 + 1 + 1 + 1. That's aggregation. **When you've read enough** heuristics and **biases research, and enough coherence** and uniqueness **proofs for** Bayesian probabilities and **expected utility**, and you've seen the "Dutch book" and "money pump" effects that penalize trying to handle uncertain outcomes any other way, **then you don't see** the **preference reversals** in the Allais Paradox **as** revealing **some** incredibly **deep moral truth** about the intrinsic value of certainty. **It just goes to show that the brain doesn't** goddamn **multiply.** The primitive, perceptual intuitions that make a choice "feel good" don't handle probabilistic pathways through time very skillfully, especially when the probabilities have been expressed symbolically rather than experienced as a frequency. So you reflect, devise more trustworthy logics, and think it through in words. When you see people insisting that no amount of money whatsoever is worth a single human life, and then driving an extra mile to save $10; or when you see people insisting that no amount of money is worth a decrement of health, and then choosing the cheapest health insurance available; then you don't think that their protestations reveal some deep truth about incommensurable utilities. Part of it, clearly, is that **primitive intuitions don't successfully diminish the emotional impact of** symbols standing for **small quantities** - anything you talk about seems like "an amount worth considering". And part of it has to do with preferring unconditional social rules to conditional social rules. Conditional rules seem weaker, seem more subject to manipulation. If there's any loophole that lets the government legally commit torture, then the government will drive a truck through that loophole. So it seems like there should be an unconditional social injunction against preferring money to life, and no "but" following it. Not even "but a thousand dollars isn't worth a 0.0000000001% probability of saving a life". Though the latter choice, of course, is revealed every time we sneeze without calling a doctor. The rhetoric of sacredness gets bonus points for seeming to express an unlimited commitment, an unconditional refusal that signals trustworthiness and refusal to compromise. So you conclude that moral rhetoric espouses qualitative distinctions, because espousing a quantitative tradeoff would sound like you were plotting to defect. On such occasions, people vigorously want to throw quantities out the window, and they get upset if you try to bring quantities back in, because quantities sound like conditions that would weaken the rule. But you don't conclude that there are actually two tiers of utility with lexical ordering. You don't conclude that there is actually an infinitely sharp moral gradient, some atom that moves a Planck distance (in our continuous physical universe) and sends a utility from 0 to infinity. You don't conclude that utilities must be expressed using hyper-real numbers. Because the lower tier would simply vanish in any equation. It would never be worth the tiniest effort to recalculate for it. All decisions would be determined by the upper tier, and all thought spent thinking about the upper tier only, if the upper tier genuinely had lexical priority. As Peter Norvig once pointed out, if Asimov's robots had strict priority for the First Law of Robotics ("A robot shall not harm a human being, nor through inaction allow a human being to come to harm") then no robot's behavior would ever show any sign of the other two Laws; there would always be some tiny First Law factor that would be sufficient to determine the decision. Whatever value is worth thinking about at all, must be worth trading off against all other values worth thinking about, because thought itself is a limited resource that must be traded off. When you reveal a value, you reveal a utility. I don't say that morality should always be simple. I've already said that the meaning of music is more than happiness alone, more than just a pleasure center lighting up. I would rather see music composed by people than by nonsentient machine learning algorithms, so that someone should have the joy of composition; I care about the journey, as well as the destination. And I am ready to hear if you tell me that the value of music is deeper, and involves more complications, than I realize - that the valuation of this one event is more complex than I know. But that's for one event. When it comes to multiplying by quantities and probabilities, complication is to be avoided - at least if you care more about the destination than the journey. **When you've reflected** on enough intuitions, **and corrected enough absurdities, you** start to **see a common denominator, a meta-principle** at work, **which one might phrase as "Shut up and multiply."** Where music is concerned, I care about the journey. When lives are at stake, I shut up and multiply. It is more important that lives be saved, than that we conform to any particular ritual in saving them. And the optimal path to that destination is governed by laws that are simple, because they are math. **And that's why I'm a utilitarian** - at least when I am doing something that is overwhelmingly more important than my own feelings about it - which is most of the time, because there are not many utilitarians, and many things left undone.

### Bostrom 2.0

#### Extinction must matter under any framework. Traditional ethics must be abandoned in the face of extinction in order to ensure that deliberation over the alternatives can continue.

Bostrom ’11 Nick **Bostrom ‘11**, “The Concept of Existential Risk”, 2011, Future of Humanity Institute, Oxford Martin School & Faculty of Philosophy http://www.existential-risk.org/concept.html

These reflections on moral uncertainty suggest an alternative, complementary way of looking at existential risk.  Let me elaborate. Our present understanding of axiology might well be confused.  We may not now know—at least not in concrete detail—what outcomes would count as a big win for humanity; we might not even yet be able to imagine the best ends of our journey.  If we are indeed profoundly uncertain about our ultimate aims, then we should recognize that there is a great *option value* in preserving—and ideally improving—our ability to recognize value and to steer the future accordingly.  Ensuring that there will be a future version of humanity with great powers and a propensity to use them wisely is plausibly the best way available to us to increase the probability that the future will contain a lot of value.  To do this, we must prevent any existential catastrophe. We thus want to reach a state in which we have (*a*) far greater intelligence, knowledge, and sounder judgment than we currently do; (*b*) far greater ability to solve global-coordination problems; (*c*) far greater technological capabilities and physical resources; and such that (*d*) our values and preferences are not corrupted in the process of getting there (but rather, if possible, improved).

### Bostrom—Parliament

#### Moral uncertainty means that we must avoid existential risk—even with low probability of util, we still have to concede high leverage issues.

Bostrom 9 (Nick Bostrom , Professor, Faculty of Philosophy & Oxford Martin School, “Moral uncertainty – towards a solution?”, January 1 2009)//Miro

It seems people are overconfident about their moral beliefs. But how should one reason and act if one acknowledges that one is uncertain about morality – not just applied ethics but fundamental moral issues? if you don't know which moral theory is correct? It doesn't seem you can simply plug your uncertainty into expected utility decision theory and crank the wheel; because many moral theories state that you should not always maximize expected utility. Even if we limit consideration to consequentialist theories, it still is hard to see how to combine them in the standard decision theoretic framework. For example, suppose you give X% probability to total utilitarianism and (100-X)% to average utilitarianism. Now an action might add 5 utils to total happiness and decrease average happiness by 2 utils. (This could happen, e.g. if you create a new happy person that is less happy than the people who already existed.) Now what do you do, for different values of X? The problem gets even more complicated if we consider not only consequentialist theories but also deontological theories, contractarian theories, virtue ethics, etc. We might even throw various meta-ethical theories into the stew: error theory, relativism, etc. I'm working on a paper on this together with my colleague Toby Ord. We have some arguments against a few possible "solutions" that we think don't work. On the positive side we have some tricks that work for a few special cases. But beyond that, the best we have managed so far is a kind of metaphor, which we don't think is literally and exactly correct, and it is a bit under-determined, but it seems to get things roughly right and it might point in the right direction: The Parliamentary Model. Suppose that you have a set of mutually exclusive moral theories, and that you assign each of these some probability. Now imagine that each of these theories gets to send some number of delegates to The Parliament. The number of delegates each theory gets to send is proportional to the probability of the theory. Then the delegates bargain with one another for support on various issues; and the Parliament reaches a decision by the delegates voting. What you should do is act according to the decisions of this imaginary Parliament. (Actually, we use an extra trick here: we imagine that the delegates act as if the Parliament's decision were a stochastic variable such that the probability of the Parliament taking action A is proportional to the fraction of votes for A. This has the effect of eliminating the artificial 50% threshold that otherwise gives a majority bloc absolute power. Yet – unbeknownst to the delegates – the Parliament always takes whatever action got the most votes: this way we avoid paying the cost of the randomization!) The idea here is that moral theories get more influence the more probable they are; yet even a relatively weak theory can still get its way on some issues that the theory think are extremely important by sacrificing its influence on other issues that other theories deem more important. For example, suppose you assign 10% probability to total utilitarianism and 90% to moral egoism (just to illustrate the principle). Then the Parliament would mostly take actions that maximize egoistic satisfaction; however it would make some concessions to utilitarianism on issues that utilitarianism thinks is especially important. In this example, the person might donate some portion of their income to existential risks research and otherwise live completely selfishly. I think there might be wisdom in this model. It avoids the dangerous and unstable extremism that would result from letting one’s current favorite moral theory completely dictate action, while still allowing the aggressive pursuit of some non-commonsensical high-leverage strategies so long as they don’t infringe too much on what other major moral theories deem centrally important.

### Cumminskey—Deon -> Util

#### Respect for human worth would justify util.

Cummiskey 90 (Cummiskey, David. Associate professor of philosophy at the University of Chicago. “Kantian Consequentiaism.” Ethics 100 (April 1990), University of Chicago. <http://www.jstor.org/stable/2381810>)

We must not obscure the issue by characterizing this type of case as the sacrifice of individuals for some abstract “social entity.” It is not a question of some persons having to bear the cost for some elusive “overall social good.” Instead, the question is whether some persons must bear the inescapable cost for the sake of other persons. Robert Nozick, for example, argues that “to use a person in this way does not sufficiently respect and take account of the fact that he is a separate person, that his is the only life he has.” But why is this not equally true of all those whom we do not save through our failure to act? By emphasizing solely the one who must bear the cost if we act, we fail to sufficiently respect and take account of the many other separate persons, each with only one life, who will bear the cost of our inaction. In such a situation, what would a conscientious Kantian agent, an agent motivated by the unconditional value of rational beings, choose? A morally good agent recognizes that the basis of all particular duties is the principle that “rational nature exists as an end in itself”. Rational nature as such is the supreme objective end of all conduct. If one truly believes that all rational beings have an equal value, then the rational solution to such a dilemma involves maximally promoting the lives and liberties of as many rational beings as possible. In order to avoid this conclusion, the non-consequentialist Kantian needs to justify agent-centered constraints. As we saw in chapter 1, however, even most Kantian deontologists recognize that agent-centered constraints require a non- value-based rationale. But we have seen that Kant’s normative theory is based on an unconditionally valuable end. How can a concern for the value of rational beings lead to a refusal to sacrifice rational beings even when this would prevent other more extensive losses of rational beings? If the moral law is based on the value of rational beings and their ends, then what is the rationale for prohibiting a moral agent from maximally promoting these two tiers of value? If I sacrifice some for the sake of others, I do not use them arbitrarily, and I do not deny the unconditional value of rational beings. Persons may have “dignity, that is, an unconditional and incomparable worth” that transcends any market value, but persons also have a fundamental equality that dictates that some must sometimes give way for the sake of others. The concept of the end-in-itself does not support the view that we may never force another to bear some cost in order to benefit others.

### Hardin—Headache

Cost-benefit analysis is feasible. Ignore any util calc indicts. **Hardin 90** writes[[5]](#footnote-5)

**One** of the **cute**r **charge**s **against util**itarianism **is that** it is irrational in the following sense. **If I take the time to calculate** the consequences of various courses of action before me, **then** I will ipso facto have chosen the course of action to take, namely, to sit and calculate, because while I am calculating the other **courses of action will cease to be open to me. It should embarrass philosophers that they have ever taken this** objection **seriously. Parallel considerations in other realms are dismissed** with eminently good sense. Lord Devlin notes, “If the reasonable man ‘worked to rule’ by perusing to the point of comprehension every form he was handed, the commercial and administrative life of the country would creep **to** a standstill.” James March and Herbert Simon **escape** the quandary of **unending calculation** by noting that often we satisfice, **we do not maximize: we stop calculating** and considering **when we find a merely adequate choice** of action. **When**, in principle, **one cannot know what is** the **best** choice, **one can nevertheless be sure that** sitting and **calculating is not the best choice.** But, one may ask, How do you know that another ten minutes of calculation would not have produced a better choice? And one can only answer, You do not. At some point the quarrel begins to sound adolescent. It is ironic that **the point** of the quarrel **is almost never at issue in practice** (as Devlin implies, **we are** almost all **too reasonable** in practice **to bring the world to a standstill**) but only in the principled discussions of academics.

### Sunstein—No action-omission

#### No action-omission decision— government must minimize the harm caused by its choices.

Sunstein 5 (Sunstein, “Is Capital Punishment Morally Required? The Relevance of Life-Life Tradeoffs”, 2005)//Miro

In our view, any effort to distinguish between acts and omissions goes wrong by overlooking the distinctive features of government as a moral agent. If correct, this point has broad implications for criminal and civil law. Whatever the general status of the act/omission distinction as a matter of moral philosophy, the distinction is least impressive when applied to government, because the most plausible underlying considerations do not apply to official actors. The most fundamental point is that unlike individuals, governments always and necessarily face a choice between or among possible policies for regulating third parties. The distinction between acts and omissions may not be intelligible in this context, and even if it is, the distinction does not make a morally relevant difference. Most generally, government is in the business of creating permissions and prohibitions. When it explicitly or implicitly authorizes private action, it is not omitting to do anything or refusing to act. Moreover, the distinction between authorized and unauthorized private action – for example, private killing – becomes obscure when government formally forbids private action but chooses a set of policy instruments that do not adequately or fully discourage it. If there is no act-omission distinction, then government is fully complicit with any harm it allows, so decisions are moral if they minimize harm. All means based and side constraint theories collapse because two violations require aggregation.

## AC

### Notes

#### The AC util shells are meant to be modified—DON’T JUST THROW THIS IN

### AC FW—Parli Util (S)

#### The standard is maximizing life.

#### Moral uncertainty means that we must avoid existential risk—even with low probability of util, we still have to concede high leverage issues.

Bostrom (Nick Bostrom , Professor, Faculty of Philosophy & Oxford Martin School, “Moral uncertainty – towards a solution?”, January 1 2009)//Miro

It seems people are overconfident about their moral beliefs. But how should one reason and act if one acknowledges that one is uncertain about morality – not just applied ethics but fundamental moral issues? if you don't know which moral theory is correct? It doesn't seem you can simply plug your uncertainty into expected utility decision theory and crank the wheel; because many moral theories state that you should not always maximize expected utility. Even if we limit consideration to consequentialist theories, it still is hard to see how to combine them in the standard decision theoretic framework. For example, suppose you give X% probability to total utilitarianism and (100-X)% to average utilitarianism. Now an action might add 5 utils to total happiness and decrease average happiness by 2 utils. (This could happen, e.g. if you create a new happy person that is less happy than the people who already existed.) Now what do you do, for different values of X? The problem gets even more complicated if we consider not only consequentialist theories but also deontological theories, contractarian theories, virtue ethics, etc. We might even throw various meta-ethical theories into the stew: error theory, relativism, etc. I'm working on a paper on this together with my colleague Toby Ord. We have some arguments against a few possible "solutions" that we think don't work. On the positive side we have some tricks that work for a few special cases. But beyond that, the best we have managed so far is a kind of metaphor, which we don't think is literally and exactly correct, and it is a bit under-determined, but it seems to get things roughly right and it might point in the right direction: The Parliamentary Model. Suppose that you have a set of mutually exclusive moral theories, and that you assign each of these some probability. Now imagine that each of these theories gets to send some number of delegates to The Parliament. The number of delegates each theory gets to send is proportional to the probability of the theory. Then the delegates bargain with one another for support on various issues; and the Parliament reaches a decision by the delegates voting. What you should do is act according to the decisions of this imaginary Parliament. (Actually, we use an extra trick here: we imagine that the delegates act as if the Parliament's decision were a stochastic variable such that the probability of the Parliament taking action A is proportional to the fraction of votes for A. This has the effect of eliminating the artificial 50% threshold that otherwise gives a majority bloc absolute power. Yet – unbeknownst to the delegates – the Parliament always takes whatever action got the most votes: this way we avoid paying the cost of the randomization!) The idea here is that moral theories get more influence the more probable they are; yet even a relatively weak theory can still get its way on some issues that the theory think are extremely important by sacrificing its influence on other issues that other theories deem more important. For example, suppose you assign 10% probability to total utilitarianism and 90% to moral egoism (just to illustrate the principle). Then the Parliament would mostly take actions that maximize egoistic satisfaction; however it would make some concessions to utilitarianism on issues that utilitarianism thinks is especially important. In this example, the person might donate some portion of their income to existential risks research and otherwise live completely selfishly. I think there might be wisdom in this model. It avoids the dangerous and unstable extremism that would result from letting one’s current favorite moral theory completely dictate action, while still allowing the aggressive pursuit of some non-commonsensical high-leverage strategies so long as they don’t infringe too much on what other major moral theories deem centrally important.

#### It is impossible for the government to do nothing—it must minimize the harm caused by its choices.

Sunstein (Sunstein, “Is Capital Punishment Morally Required? The Relevance of Life-Life Tradeoffs”, 2005)//Miro

In our view, any effort to distinguish between acts and omissions goes wrong by overlooking the distinctive features of government as a moral agent. If correct, this point has broad implications for criminal and civil law. Whatever the general status of the act/omission distinction as a matter of moral philosophy, the distinction is least impressive when applied to government, because the most plausible underlying considerations do not apply to official actors. The most fundamental point is that unlike individuals, governments always and necessarily face a choice between or among possible policies for regulating third parties. The distinction between acts and omissions may not be intelligible in this context, and even if it is, the distinction does not make a morally relevant difference. Most generally, government is in the business of creating permissions and prohibitions. When it explicitly or implicitly authorizes private action, it is not omitting to do anything or refusing to act. Moreover, the distinction between authorized and unauthorized private action – for example, private killing – becomes obscure when government formally forbids private action but chooses a set of policy instruments that do not adequately or fully discourage it. If there is no act-omission distinction, then government is fully complicit with any harm it allows, so decisions are moral if they minimize harm. All means based and side constraint theories collapse because two violations require aggregation.

#### Two implications

#### Takes out standard indicts—side-constraints are inevitably violated

#### No action/omission distinction makes questions of skep and permissiveness irrelevant.

1. Ezio Di Nucci (Associate Professor of Medical Ethics at the University of Copenhagen). “Eight Arguments against Double Effect.” Proceedings of the XXIII. Kongress der Deutschen Gesellschaft für Philosophie (forthcoming). 2014. https://www.academia.edu/7841331/Eight\_Arguments\_against\_Double\_Effect [↑](#footnote-ref-1)
2. Derek Parfit, Reasons and Persons (Oxford: Clarendon, 1984). [↑](#footnote-ref-2)
3. Shoemaker, David (Dept of Philosophy, U Memphis). “Utilitarianism and Personal Identity.” *The Journal of Value Inquiry* 33: 183–199, 1999. <http://www.csun.edu/~ds56723/jvipaper.pdf> [↑](#footnote-ref-3)
4. Laura Valentini. “Kant, Ripstein and the Circle of Freedom: A Critical Note.” European Journal of Philosophy. 2012. [↑](#footnote-ref-4)
5. Hardin, Russell (Helen Gould Shepard Professor in the Social Sciences @ NYU). May 1990. Morality within the Limits of Reason. University Of Chicago Press. pp. 4. ISBN 978-0226316208. [↑](#footnote-ref-5)