A Bestiary of Utility Monsters

Walter Barta

University of Houston

(First Draft February 2021)

(Revised for the NMTX Philosophy Conference 2024)

Introduction

The concept of the Utility Monster offers an influential critique of Utilitarian theories, forcing us to consider different theoretical fixes to escape monstrous implications (Nozick, 1999, pp. 26-53; Kennard, 2015, p. 322). However, many different breeds, a whole bestiary, of Utility Monsters are identifiable, and each breed reveals something slightly different about what we find monstrous. When dissected in depth, we observe that some breeds are probably acceptable, whereas other breeds are indeed monstrous, though perhaps for slightly different reasons than Nozick thought. By breaking these taxonomies down, thus revealing strengths and weaknesses of these different breeds, exhausting the conceptual space occupied by such monsters, we may see more clearly how the vicious versions can be dispensed with under reasonable assumptions.

The Original Monster and its Would-be Defeaters

The original formulation of the Utility Monster was a critique of Total Utilitarianism, the view that we should maximize the Total Utility, the utility of the population, arguments summarized as the "greatest good for the greatest number" (Bentham, 2019, p. 7). Nozick's "utility monsters"[1] are thought experiments showing that Total Utilitarianism, Average Utilitarianism, and other theories[2] of utility are consistent with optimal outcomes in which a singular over-consuming person experiences most or all of the values and the rest of the population experiences little or none (Nozick, 1999, p. 41). What seems monstrous about these implications is that altruistic Utilitarianism seems to collapse into (and ethically justify) pure egoism for the monster itself, and suicidal masochism for every other person. But, to add insult to injury, by definition, feeding the Utility Monster is obligatory; to not sacrifice oneself to the Utility Monster would be normatively monstrous. This is because, by definition, Utility Monsters have their Personal Optimality (bliss point), the optimal utility state of specific persons, coincide with the Total Optimality, the optimal utility state of the population (Binger et al., 1998, p. 113). So, per the thought experiment, if Utilitarianism is true and Utility Monsters exist, then we must prioritize

[1] Nozick's original description: "Utilitarian theory is embarrassed by the possibility of utility monsters who get enormously greater sums of utility from any sacrifice of others than these others lose ... the theory seems to require that we all be sacrificed in the monster's maw, in order to increase total utility" (Nozick, 1999, p. 41).

[2] Nozick's other major target of criticism is John Rawls' maximin principle, "the greatest benefit of the least advantaged," which Nozick also believes is vulnerable to Utility Monsters, if the least advantaged behave as unconditionally demanding free-riders at the expense of every other contributing citizen (Rawls, 1971, p. 302).

the monsters at the expense of everything else, including other Personal Utilities, the utilities of specific persons. Thus, the Utility Monster shows that certain versions of Utilitarianism have counterintuitive consequences, and so Utilitarians should only accept versions of their normative system that exclude Utility Monsters—but problematically, regular unbounded Utilitarianism doesn't. Prima facie, this seems like a problem, so how do we solve it?

A first solution: we may simply bite the bullet and accept that we are normatively required to prioritize the monster (Fisher, 2020). Although this requires sacrificing ourselves, that is just what duty requires. Biting the bullet though seems far too counterintuitively demanding, at least for a first pass normative intuition, so we should not accept this as our default position and perhaps only accept it after alternatives are exhausted.

A second solution: we may, as Nozick does, reject personal "sacrifice" as normatively required by setting up a "moral side constraint", thus rejecting versions of utilitarianism that require the monster (Nozick, 1999, p. 32-33). This seems appealing, but this fails to account for some situations in which unwillingness to sacrifice would seem monstrous, for example, if one does not save a girl from drowning in a shallow pond because one does not what to get we (Singer, 1973). So, Nozick's solution seems to fail because it fails to accommodate such cases.

A third solution: we may reject possibility of "interpersonal utility comparisons",[3] thus rejecting the assumption that the Monster's personal utility is comparable with anyone else's personal utility (Hausman, 1995, pp. 475-76). This seems appealing because some values really seem unmeasurable and incommensurable, but the approach does not seem to fully address the problem. First, some objective values do indeed seem comparable and monsters can arise in these domains. Second, we can imagine artificially designed agents with publicly accessible mind (perhaps a Momus Glass is installed in their head), thus rendering even their subjective values transparent and measurable. Third, whether values are comparable or not, practical action often seems to demand acting as though they are comparable out of necessity so as to be able to discriminate between options at all. Thus, the incomparability solution seems inadequate to all cases.

A fourth solution: we may conclude, as Parfit does, that Utility Monsters of Nozick's description do not actually exist. Parfit points out they would need to be "millions" of times happier than any other humans,[4] which seems to describe no actual humans (Parfit, 1984, p. 388-89). However, what Parfit's objection misses is that, even if Utility Monsters are not actual but merely possible,

---

[3] Hausman's objection reads: "…comparisons of utility differences, are arbitrary for the simple reason (which has nothing to do with interpersonal comparisons) that utility differences are arbitrary. With only ordinal utilities . . . Nozick's "utility monster" . . . is impossible . . . cardinality doesn't resurrect the monster either." (Hausman, 1995, p. 475-76).

[4] Parfit's objection reads: "For this to be true, this Monster's quality of life must be *millions* of times as high as that of anyone we know. Can we imagine this? Think of the life of the luckiest person that you know, and ask what a life would have to be like in order to be a million times as much worth living. The qualitative gap between such a life and ours, at its best, must resemble the gap between ours, at its best, and the life of those creatures who are barely conscious—such as, if they *are* conscious, Plato's 'contented oysters'. It seems a fair reply that we cannot imagine, even in the dimmest way, the life of this Utility Monster. And this casts doubt on the force of the example." (Parfit, 1984, p. 388-89).

then we may be obliged to consider them anyway, as a matter of theory. Furthermore, given that it may be (and probably is) possible to create artificial sentient agents, we may someday be able to (and thus be obliged to) create a Utility Monster to sacrifice ourselves to (Frank, 2000).

So, if we are unsatisfied with all of these options we may need to further dissect the Utility Monster (and its ilk) in order to identify what really bothers us about such beings.

Universal Utility Monsters

The original Utility Monster concept was Universal in the sense that it stipulated that "any" added goods (here used to mean "commodities") for "all" persons give themselves to the monster (Nozick, 1999, p. 41). We can give this type of Utility Monster a more rigorous description here:

Definition: a person that has greater marginal utility than all other persons in all situations. (We have prefaced this "Universal" because it is such for all domains, which will contrast other forms.)

Mathematics: one arises when a person has a high-rate linear commodity-utility function (high constant marginal utility).

We have also formulated these mathematical features as equations in Appendix A and as graphs in Appendix B. The figures show utility and marginal utility across a domain of commodities, a utility monster of a specific type compared against a normal agent with diminishing marginal utility (Mises, 1998).

Consequences: Other persons are never prioritized; the monster is prioritized indefinitely.

Real World Example: a human with a stoma implant in their stomach that can relieve food storage from the stomach as it accumulates is able to extract pleasure from food without their pleasure ever having to be bounded and diminished by feelings of satiety.

Good Features: the Universal Utility Monster does not seem to have any good features, besides for the good feature implicit to all Utility Monsters, which is that increasing the monster's personal utility maximally increases the total utility of the population.

Bad Features:

> **Incidental Inequality**: The monsters are prioritized over other persons under certain circumstances in which they stand to gain greater utility.
>
> **Zero Personal Utility**: None of the persons are ever entitled to achieve personal utility above zero, excepting the Utility Monster itself, which is itself entitled to the greatest possible utility within the domain of possibility, but less than its optimal (infinite) utility at infinity. The Total Optimal is never the personal optimal.
>
> **Indefinite Commoditys Requirement**: The marginal utility of commodities persists indefinitely, such that an indefinitely increasing quantity of commodities will always increase the total utility.
>
> **Indefinite Utility Requirement**: The growth of utility can persist indefinitely, such that total utility can never quite achieve Total Optimality.

Partial Utility Monsters

However, we can stipulate that there may be other kinds of Utility Monsters, non-universal (partial) Utility Monsters: kinds of utility monster that are monstrous only for some domains. These Partial Utility Monsters may themselves come in many different varieties depending upon where the threshold of monstrosity falls within the domain: there are as many partial kinds as there are kinds of asymptotic behavior for utility in the domain of commodities. Some thresholds may be preferable to others. Notably, we will suggest that there are at least four different kinds of identifiable Utility Monster, which we will discuss in turn—Sub-optimal, Para-Optimal, Optimal, and Supra-Optimal. Dissecting out these variant breeds will help us exhaust the Cartesian space that can be occupied by utility/commodity relationships, then diagnose what specifically we find so monstrous about Utility Monsters, and finally prescribe solutions.

Sub-Optimal:

Definition: a person that has greater marginal utility than all other persons in some personally sub-optimal situations.

Mathematics: one arises when a person has a low-rate linear commodity-utility function (low constant marginal utility).

Consequences: In the presence of such a monster, other persons are briefly prioritized, until some diminished marginal utility, beneath their optimum, at which the monster's marginal utility exceeds their marginal utilities, after which the monster is prioritized indefinitely, for an infinite quantity of commodities.

Real World Example: a king permits his peasants to eat food but never enough to be entirely satisfied, and the king takes the rest of their food for his private court of nobles who gorge themselves ceaselessly to their indefinitely pleasure (Hobbes, Chapter XV).

Good Features:

> **Positive Personal Utility**: none of the other persons are ever condemned to subsist at zero or subzero utility. Rather, persons are entitled to some sub-optimal but positive level of personal utility. The Total Optimal is never personally miserable.

Bad Features:

> **Incidental Inequality**: Same as above.
>
> **Sub-optimal Personal Utility**: Similar to above.
>
> (None of the persons are ever entitled to achieve their personal optimality (bliss point), including for the monster itself, which is itself entitled to the greatest possible utility within the domain of possibility, but far less than its optimal (infinite) utility at infinity. The Total Optimal is never the personal optimal.)
>
> **Indefinite Commodity Requirement**: Same as above.
>
> **Indefinite Utility Requirement**: Same as above

Para-Optimal:

Definition: a person that has greater marginal utility than all other persons in some personally para-optimal (near optimal) situations.

Mathematics: one arises when a person has an asymptotic commodity-utility function approaching some positive horizontal limit (decaying marginal utility).

Consequences: In the presence of such a monster, the monsters are prioritized until their marginal utility diminishes to equal some other persons' marginal utility, at which point those other persons are prioritized until their marginal utility diminishes to equal some other persons' marginal utility, indefinitely, always nearly but never quite optimal, at which point their marginal utilities are neck-and-neck with the monster's marginal utility indefinitely, without ever quite reaching their optimum, for an infinite quantity of commodities.

Real World Example: again, a king taxes his citizens progressively based on their satisfaction level, always taking more food the closer they get to total satisfaction, and takes the rest for his private court of nobles who do not get much extra pleasure from the extra food.

Good Features:

> **Near-Optimal Personal Utility**: none of the other persons are ever condemned to subsist at low levels of personal utility. Rather, persons are entitled to some nearly optimal level of personal utility. The Total Optimal is personally near-optimal.

Bad Features:

> **Incidental Inequality**: Same as above.
>
> **Sub-optimal Personal Utility**: Similar to above.
>
> **(**Though, this is better than for the Sub-optimal case because personal utility is near-optimal.)
>
> **Indefinite Commodity Requirement**: Similar to above.
>
> **(**Though, this is worse than for the Sub-optimal case because the Indefinite Commodity required offer vanishingly small marginal utilities.)

Optimal:

Definition: a person that has greater marginal utility than all other persons in some personally optimal situations.

Mathematics: one arises when a person has a negative polynomial commodity-utility function (decreasing marginal utility); or a constant (horizontal) positive commodity-utility function.

Consequences: In the presence of such a monster, the monster is prioritized until their marginal utility diminishes to equal some other persons' marginal utility, at which point those other persons are prioritized until their marginal utility diminishes to equal some other persons' marginal utility, until every person has reached their respective optimum at some finite quantity of commodities.

Real World Example: a race of giant humans governs a race of miniature humans, and the giant humans extract the same amount of nutritive value from every pound of food as two miniature humans, although both races can reach satiety given enough food is available to them.

Good Features:

> **Optimal Personal Utility**: persons are entitled to the optimal level of personal utility. The Total Optimal is personally optimal.

Finite Commodity Requirement: The marginal utility of commodities persists until some finite quantity, such that an indefinitely increasing quantity of commodities will not indefinitely increase the total utility.

Bad Features:

Incidental Inequality: same as above.

Supra-Optimal:

Definition: a person that has greater marginal utility than all other persons in some personally supra-optimal situations.

Mathematics: A Supra-optimal Utility Monster arises when a person has a positive polynomial commodity-utility function (increasing marginal utility); or a constant (horizontal) positive commodity-utility function.

Consequences: In the presence of such a monster, other persons are prioritized, until the diminished marginal utility at their optimums, after which point the monster is prioritized indefinitely.

Real World Example: a private club makes sure that every single existing member is completely satisfied, but then pulls a profit from its members indefinitely after. Furthermore, the club does not allow new members to join, and instead treats these new members as non-persons who are not entitled to anything.

Good Features:

Optimal Personal Utility: Same as above.

Bad Features:

Zero Personal Utility (For New Persons): Similar to above.

(Though, only once the marginal utility of the monster increases sufficiently, the marginal utility of the monster may exceed the marginal utility of any new persons added to the domain of consideration, such that the monster never permits the entry of new person into consideration.)

Indefinite Commodity Requirement: Similar to above.

(Though this is better than the Para-Optimal case because the Indefinite Commodity required offer increasing marginal utilities.)

Indefinite Utility Requirement: Same as above

Monstrosities and their Defeaters

In what follows, we will discuss the four major problems identified regarding Utility Monsters and will attempt to resolve each problem respectively. We will show that the first three seeming problems are intuitively acceptable, and we will show that the final problem can be excluded by properly specifying the normative domain.

Table 1: Utility Monster Breeds and their Problems

	INCIDENTAL INEQUALITY	SUBOPTIMAL PERSONAL UTILITY	INDEFINITE COMMODITY REQUIRED	INDEFINITE UTILITY REQUIRED
UNIVERSAL	TOTAL	ZERO	TRUE	TRUE
SUPEROPTIMAL	TRUE	NEW PEOPLE	TRUE	TRUE
SUBOPTIMAL	TRUE	POSITIVE	TRUE	TRUE
PARAOPTIMAL	TRUE	BARELY	TRUE	ALMOST
OPTIMAL	TRUE			

Incidental Inequality

The Partial Optimal Utility Monster has the downside of obligating situations of Incidental Inequality. In any given instance of added commodities, we might be obligated to give more of that commodity to Person A than to Person B until the marginal utility of Person A and B balances.

However, upon reflection, Incidental Inequality does not seem like a problem. Indeed, we routinely encounter situations of Incidental Inequality, such that they are trivial. On the banal end of situations, a slightly taller person might require slightly more food for sustenance than a slightly shorter person. On the emergency end of situations,[5] a dying person may require more urgent and stringent medical attention than a person suffering a minor injury (Singer, 1972). These temporary inequalities do not offend our intuitions; in fact, rejecting these cases on the grounds of unfairness would seem monstrous itself. Indeed, when reflecting upon the nature of action in the world, we can observe rather straightforwardly that differential action is required as a condition of doing anything, such that temporary inequalities of some duration and magnitude would have to be suffered even by the most egalitarian principles, if only just to avoid paralysis for long enough to get anything done (lest we find ourselves in the position of a Buridan's Ass-Altruist: someone unable to choose between which charity to give to who ends up forgoing the opportunity to give) (Aristotle, 2020, 295b). Generalized equality does not imply instantaneous equality at every point in spacetime. So, not only are personal Incidental sacrifices understandable, but they are also necessary in the limit case if only to avoid Buridan's Ass styled dilemmas.

Notably though, even if we accept incidental inequality, it seems that the Universal Utility Monster requires *total inequality*, which might remain uniquely acceptable.

Sub-Optimal Personal Utility

The Partial Sub-optimal and Para-Optimal Utility Monster have the downside of obligating situations of Sub-optimal Personal Utility. For any given case of Total Optimal Utility, any given

[5] Some such emergency cases seem so egregious that it would seem monstrous for Nozick to object to them on the grounds of libertarian constraint (Nozick, 1999, p. 45).

person might not be permitted optimal personal utility. For instance, at a given Total Optimum, Person A might be obligated to sustain a sub-optimal personal utility.

However, upon reflection, although Sub-Optimal Personal Utility seems like a personal problem, as a consideration of the Total Utility it does not seem like a problem. Indeed, we can easily imagine that some Total Optimums may permit and even require Personal Suboptimums. On the banal end of such situations, given finite amounts of food, a person might be obliged to not eat as much as they want so that they can share with another person. On the emergency end of such situations, given certain dangers, some number of persons might have to sacrifice wellbeing to protect the optimums of the rest of the population. These personal sacrifices to the total wellbeing need not offend our intuitions; in fact, while these circumstances may be lamentable, they may not be preventable given material constraints in the world (Sidgwick, 1962, p. 162). Indeed, reflecting upon the nature of material tradeoffs,[6] we find no guarantee that all optimums will coincide for every person in every situation (Pareto, 1971). The default assumption should be that optimums can sometimes conflict (Miettinen, 1999, pg. 5). (Benefiting you may necessarily cost me, and vice versa.) Total Optimality does not imply personal optimality (bliss point) for every person in the population. So, not only are personal suboptimums understandable, they are also necessary in the limit case as a material constraint on material beings.

Notably though, even if we accept suboptimal personal utility, it seems that the Universal Utility Monster requires *zero person utility*, which might remain uniquely acceptable.

Indefinite Commodity Requirements

Every Utility Monster (except the Partial Optimal) has the downside of requiring indefinite usage of commodities. The problem more rigorously defined is that for any given case of monotonically increasing utility in a domain of increasing quantities of commodities, there must always be some utility higher than the previous utility, which implies that there must always be some greater set of commodities to pursue beyond the previous sets of commodities. For instance, for a given population of indefinitely required commodities, Person A and B should not be satisfied with x commodities because they can be satisfied with x+1, and should not be satisfied with x+1 because they can be satisfied with x+2, etc.

Like the previous two problems, Indefinite Commodity Requirements do not seem obviously bad. If one of something is good, ceteris paribus, isn't two of something better? In finite ranges, this is reasonable. However, unlike the two previous problems, extrapolated to the extreme, the requirement becomes a problem, as the logical conclusion of an indefinite transitive sequence of values is that the highest value must be infinite. For instance, Person A and B should not be satisfied with x+2 because they can be satisfied with ∞. Indefinite Commodity Requirements end up implying Infinite Commodities Requirements. This is a problem because acquiring infinite commodities is a material impossibility.

This argument ad infinitum reveals the weakness of Indefinite Commodity Requirements, that we might defeat it once and for all. Namely, in any given circumstance for any given set of persons, infinite commodities requirements should be excluded from normative considerations as

[6] Nozick would surely be in agreement here, as the background of material tradeoffs is implied by his analysis (Nozick, 1999, p. 33).

contradictions: an impossible necessity. One can see this as a special instance of "ought implies can" (Kant, 1793, p. 94). Because we cannot consume infinite material commodities, we ought not consume infinite material commodities, and so we ought to reject any normative systems that end up implying that we are required to do so, as do versions of Utilitarianism that permit of Utility Monsters. Accepting this principle eliminates all Utility Monsters (excepting the Partial Optimal ones) from normative consideration and thus all theories friendly to them, including any versions of infinite-growth Utilitarianism that are not bounded by the possibilities of the material world. Thus, contra-Nozick, we propose not a "moral side constraint", Kantian "inviolability",[7] but an alethic side constraint, Kantian impossibility, as a defeater of the Utility Monster (Nozick, 1999, p. 32-33). This alethic defeater is distinct from Parfit's objection because whereas Parfit merely rejects the actuality of such monsters we have taken seriously the possibility of such monsters and then rejected that possibility.

However, all this being said, this alethic constraint can itself be rejected on the grounds that we can redefine the Utility Monster as a being with the imperative to consume all commodities *possible* (not indefinite/infinite commodities), thus stipulating some "possible" upper bound on the Indefinite Commodity Requirement. Such a Utility Monster, possible by definition, can still fill the universe with its own utility and the stop at the edge of the universe. In this framing though, the "possible commodities requirement" does not seem like a problem, since it merely imagines using all available resources to increase the utility of the universe, requiring no ethical absurdities, thus resolving our concerns. Though this solution still permits Para-Optimal Utility Monsters, we these Para-Optimal Utility Monsters do not pose any further problem that must be resolved—thus we can deign them normatively acceptable.

Indefinite Personal Utility Requirements

Finally, every Utility Monster (except the Partial Optimal and Partial Para-Optimal) has the downside of requiring indefinite utility. The problem is that there must always be some utility higher than the previous utility. For instance, for a given population of indefinitely required utility, Person A and B should not be satisfied with x utility because they can acquire x+1, and should not be satisfied with x+1 because they can acquire x+2, etc.

Like with the Indefinite Commodity Requirement, the Indefinite Utility Requirement does not seem obviously bad. If utility is good, ceteris paribus, isn't more utility better? However, like with commodities, extrapolated to the extreme, the requirement becomes infinite. Infinite utility ∞ is better than x+2 utility. Indefinite Utility Requirements end up implying Infinite Utility Requirements. This is a problem because, like acquiring infinite commodities is a material impossibility, acquiring infinite utility seems to be a psychological impossibility.

However, there are reasonable defeaters to consider to against the infinite utility requirement. Namely, we might propose a set of reasonable psychological constraints. To do this, we might observe that most Utility Monsters are necessarily psychologically exotic, perhaps even

[7] On Nozick's account, the problem arises when one is normatively required to "sacrifice" some persons to other persons, and his solution is to categorically exclude such occasions; thus, he departs greatly (which is by-and-large the content of *Anarchy, State, and Utopia*) with our interpretation of the problem and our proposal for a solution (Nozick, 1999, p. 45).

psychologically impossible. As a comparison class we can consider a normal psychological agent as one with a diminishing marginal utility function (Mises, 1998). Only the Optimal Utility Monster has a (normal) diminishing marginal utility function. In contrast, a Universal or Sub-Optimal Utility Monster will have a constant marginal utility function, implying a mind that oddly has no preference-ordering that will approach an infinite utility the more it consumes. A Supra-Optimal Utility Monster will have accelerating marginal utility function, implying a mind that oddly eventually approaches an infinite marginal utility. A Para-Optimal Utility Monster will have an always positive marginal utility function, implying a mind that oddly has no negative marginal utility. If we can exclude some or all of these exotic psychologies then some or all Utility Monsters can be ruled out. It seems like Universal/Sub-Optimal monster minds require static psychologies, unaffected by any differential consumption, which seems impossible in a dynamic world; as well as infinite positive utility values, which seem impossible in a finite mind. It seems like a Supra-Optimal monster minds require psychological preference sets that include infinite positive marginal utility values, which seem impossible in a finite mind. It seems like Para-Optimal Monster minds require psychological preference sets that are positive throughout their entire domain, without any aversions whatsoever, which seems perhaps possible but wildly maladaptive since it seems unlikely that psychologies without aversions would ever have survived their environmental dangers and evolved amidst Darwinian pressures (Darwin, 1859). However, if we are allowing for exotic artificial non-Darwinian psychologies, then Para-Optimal Monster minds remain possible, though this does not seem to entail any other normative problems and allows other persons near-optimal personal utility and therefore seems acceptable.

Thus, by process of elimination, only acceptably psychologically finite kinds of Utility Monsters, the ones with Optimal (and perhaps non-Darwinian Para-Optimal) monster minds are undefeated.

Total Inequality and Zero Personal Utility

Notably, the *total inequality* and *zero person utility* problems of the Universal Utility problems have still not been resolved.

But, if we still feel that we must resolve the *total inequality* and *zero person utility* problems of the Universal Utility Monster, with Nozick, we might impose some "moral side constraint". The solution, per Nozick, is to strongly stipulate that the individual is *always inviolable* (Nozick, 1999, p. 45). However, we have shown this solution to itself be monstrous because it permits us to completely callously ignore the sufferings of others (see the "sub-optimal personal utility" section above). So avoid this equally monstrous conclusion, contra-Nozick, we would want to weaken the constraint and thus may propose instead to weakly stipulate that the individual is *not always violable*, a subtle but important distinction. Relative to others, it is not unconditional inviolability but conditional violability; it is not that we *cannot ever* use persons as means to our ends, but that we *cannot always* use persons as means to our ends. The Universal Utility Monster feels so monstrous because it requires unconditional violations in an domain of Indefinite Commodity requirements, which becomes acceptable when the commodities requirements are made definite and the violations conditional. If we accept a "not always violable" constraint, then Universal Utility Monsters are defeated. However, adding a moral side constraint may seem unjustified to some, so this solution may be criticized on the grounds that we are merely gerrymandering our normative domain to exclude counterintuitive features.

Furthermore, we can argue the *total inequality* and *zero person utility* problems need not be resolved, since excluding exotic psychologies excludes the Universal Utility Monster and thus resolves the *total inequality* and *zero person utility* problems vicariously. The moral side constraint is unnecessary if a psychological constraint is assumed.

Table 2: Utility Monster Breeds and their Defeaters

	INCIDENTAL INEQUALITY	**SUBOPTIMAL PERSONAL UTILITY**	**INDEFINITE COMMODITY REQUIRED**	**INDEFINITE UTILITY REQUIRED**
UNIVERSAL	TOTAL	ZERO	OK	IMPOSSIBLE
SUPEROPTIMAL	OK	OK	OK	IMPOSSIBLE
SUBOPTIMAL	OK	OK	OK	IMPOSSIBLE
PARAOPTIMAL	OK	OK	OK	UNFIT
OPTIMAL	OK			

Monsters' Treasures

A first upshot: by dispensing with infinite utilities, we have constrained the possible utility functions (the mathematics by which we assign values to persons and commodities) consistent with our non-monstrous intuitions, narrowed from Bentham's unbounded types down to a narrow acceptable range: the utility function must not be monotonically increasing (Broome, 2004). Although this result may seem counterintuitive, it is consistent with many of the intuitions that we already have: to say that values are not monotonically increasing can be construed as a more technically rigorous statement of Aristotle's ethic of moderation against greed (*pleonexia*): virtuous means and vicious extremes (Aristotle, 2011).

A second upshot: this excludes a wide range of strange utility functions as unacceptable. First, as a matter of public policy, we should reject as monstrous any normative standard that assumes or implies infinite values (for example Parfit's "Repugnant Conclusion") (Parfit, 2004). This might include investors demanding infinite growth models of the economy; this might also include free-riders demanding bottomless safety-nets. Second, given that future generations may design artificial agents and specify the utility functions governing those agents, we should decline to view any agents as demanding monotonic increasing utility functions (Fisher, 2020). These might include institutions, like companies with high profit margins; these might also include machines, like algorithms bent upon perpetual optimization. As has been shown, these types of monster are technically impossible to embody in finite material.

A third upshot: all this being said, Optimal (and Para-Optimal) Utility Monsters remain possible, though this is perhaps acceptable. Most Optimal Utility Monsters will simply be sentient creatures with different marginal utility curves (e.g. persons with differing eudemonic set points) of variation that should be expected of any population (Bartels et al., 2009). Some artificial Optimal Utility Monsters may exist with exceptionally economical personal optimalities (bliss points). Arguably though, these monsters are not monstrous at all, in two respects: 1) they are more efficient commodity-utilizers; 2) they do not require the universal deprivation of others,

only situational deprivation when marginal utility favors. Perhaps some of the consequences of the Optimal Utility Monster may appear prima facie odd. For instance, a world of dwarves can be better than a world of giants, ceteris paribus; thus, an odd corollary may follow, a preference for the most efficient possible eudemonic creature (Shulman, 2012). However, this corollary is somewhat intuitive, if we think of the preference is justified by dwarves being more satisfiable creatures; or at least it is less counterintuitive than a dragon that is justified in eating everyone, so it seems an acceptable concession. If having such a preference is unacceptable, we would need to get rid of it by re-appealing to incommensurability (Hausman, 1995), or by appealing to some principle of equality that accommodates all persons involved satisfactorily (Sher 2014), valuing giants and dwarfs on par.

Conclusion

So, we have described how Total Utilitarianism is vulnerable to a bestiary of two different genera of Utility Monster (Universal and Partial), and that Partial Utility Monsters come in several different species, including: Sub-optimal, Para-Optimal, Optimal, and Supra-Optimal. We have shown that these breeds of Utility Monster impose four major problems: Incidental Inequality, Sub-Optimal Personal Utility, and Indefinite Commodity/Utility Requirements. However, all three of these problems are defeated by some reasonable assumptions—we ignore these assumptions at our peril, as that way there be monsters.

Appendix A: Equations

The following formulas describe utility monsters and persons at utility (u), in terms of commodities added to the population (x), for persons (i), given the existence one utility monster ($i = m$), some commodity level of personal optimum ($x_{i,opt}$), and some arbitrary commodity level less than personal optimum ($x_{i,sub}$).

A.1: Universal

$$\forall x \forall i \left(\frac{du_m}{dx}(x) > \frac{du_i}{dx}(x) \right)$$

A.2.a: Sub-optimal

$$\forall x \forall i \left(\frac{du_m}{dx}(x_{i,sub} < x < x_{i,opt}) > \frac{du_i}{dx}(x_{i,sub} < x < x_{i,opt}) \right)$$

A.2.b: Para-Optimal

$$\forall x \forall i \left(\frac{du_m}{dx}(x \approx x_{i,opt}) > \frac{du_i}{dx}(x \approx x_{i,opt}) \right)$$

A.2.c: Optimal

$$\exists x \forall i \left(\frac{du_m}{dx}(x) > \frac{du_i}{dx}(x) \right)$$

A.2.d: Supra-Optimal

$$\forall x \forall i \left(\frac{du_m}{dx}(x > x_{i,opt}) > \frac{du_i}{dx}(x > x_{i,opt}) \right)$$

Appendix B: Graphs

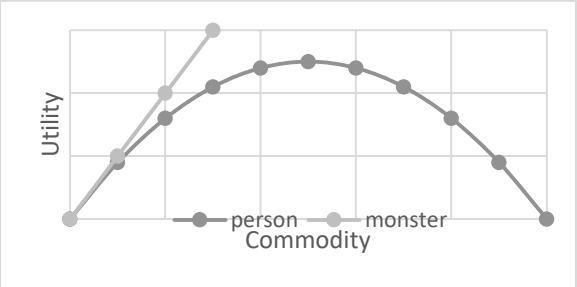

Figure 1: Universal Utility Monster

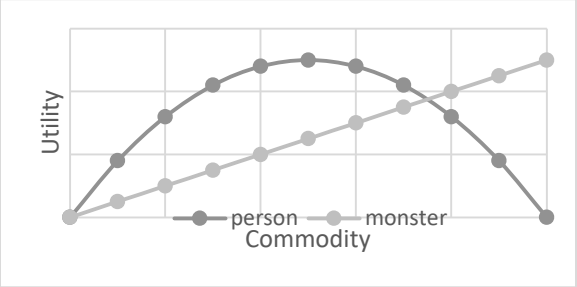

Figure 2: Sub-Optimal Utility Monster

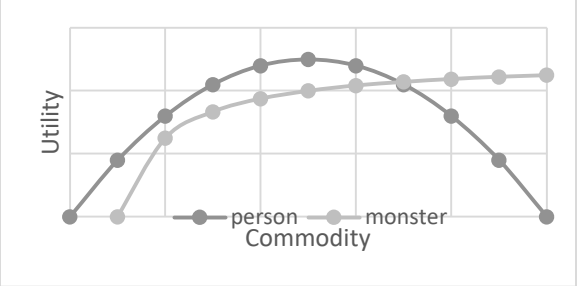

Figure 3: Para-Optimal Utility Monster

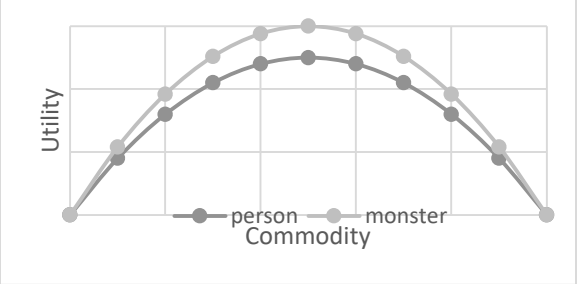

Figure 4: Optimal Utility Monster

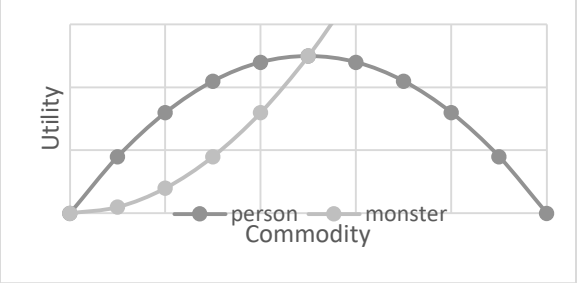

Figure 5: Supra-Optimal Utility Monster

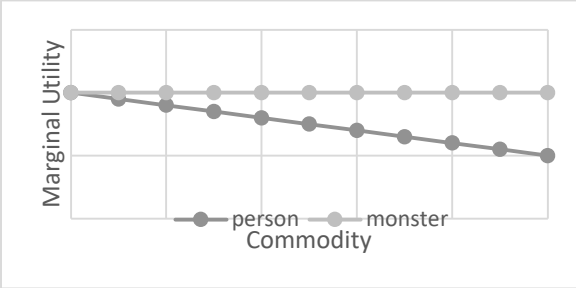

Figure 6: Universal Utility Monster Marginal Utility

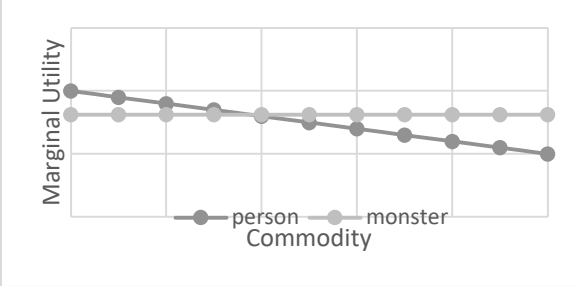

Figure 7: Sub-Optimal Utility Monster Marginal Utility

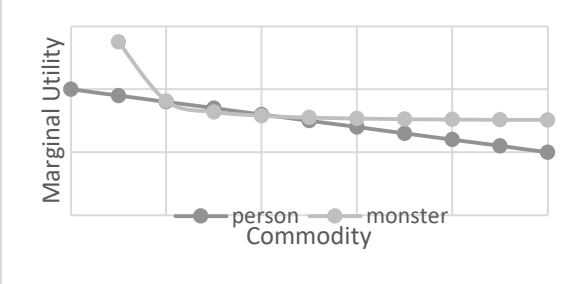

Figure 8: Para-Optimal Utility Monster Marginal Utility

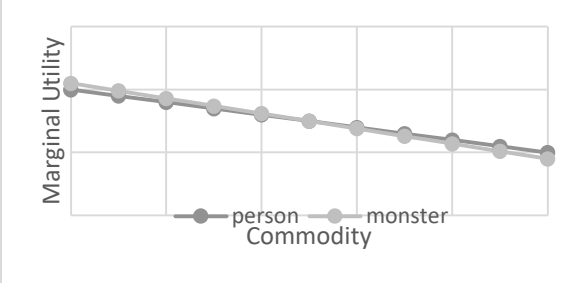

Figure 9: Optimal Utility Monster Marginal Utility

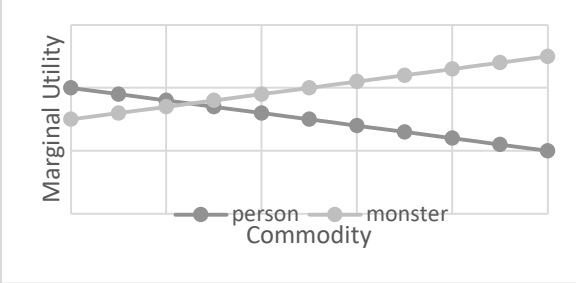

Figure 10: Supra-Optimal Utility Monster Marginal Utility

Works Cited

Aristotle. *De Caelo*. Translated by C. D. C. Reeve, Hackett, 2020, Indianapolis.

Aristotle. *Nicomachean Ethics.* Translated by Bartlett, Robert C. and Susan Collins, Chicago: University of Chicago Press, 2011, Chicago.

Bartels, M., & Boomsma, D. I. (2009). Born to be happy? The etiology of subjective well-being. *Behavior Genetics*, *39*(6), 605–615.

Bentham, Jeremy. *An Introduction to the Principles of Morals and Legislation*. Anodos Books, 2019.

Binger, B.; Hoffman, E. (1998). Microeconomics with Calculus (2nd ed.). Addison-Wesley.

Broome, John. *Weighing Lives*, Oxford University Press, 2004, Oxford.

Darwin, Charles & Kebler, L. *On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life*. London: J. Murray, 1859. [Pdf] Retrieved from the Library of Congress, https://www.loc.gov/item/06017473/.

Fisher, Richard. "The intelligent monster that you should let eat you." BBC News, 13 November 2020*, https://www.bbc.com/future/article/20201111-philosophy-of-utility-monsters-and-artificial-intelligence.* Retrieved 12 February 2021.

Frank, Robert H. (June 2000). "Why Is Cost-Benefit Analysis so Controversial?" *The Journal of Legal Studies*, Volume 29, Number S2, June 2000, pp. 913-930.

Hausman, Daniel. "The Impossibility of Interpersonal Utility Comparisons." *Mind*, Volume 104, Issue 415, July 1995, pp. 473–490.

Hobbes, Thomas. *Leviathan.* Edited by Rogers, G. A. J., Schuhmann, Karl. London: Bloomsbury Publishing, 2006, London.

Kant, Immanuel. *Religion within the Boundaries of Mere Reason*. Edited and Translated by Wood, Allen and George Di Giovanni, Cambridge University Press, 1998, Cambridge.

Kennard, Frederick. *Thought Experiments: Popular Thought Experiments in Philosophy, Physics, Ethics, Computer Science & Mathematics* (1st ed.). AMF, 2015.

Miettinen, Kaisa. *Nonlinear Multiobjective Optimization*. Springer, 1999.

Mises, Ludwig Von. *Human Action: A Treatise on Economics*. Auburn, Alabama, The Ludwig von Mises Institute, 1998.

Nozick, Robert. *Anarchy, State, and Utopia*. Blackwell Publishers Ltd., 1999, Oxford.

Pareto, Vilfredo. *Manual of Political Economy*. Translated by Ann Schwier, Augustus M. Kelley Publishers, 1971, New York.

Parfit, Derek. *Reasons and Persons*. Clarendon Press, 1984, Oxford.

Parfit, Derek. "Overpopulation and the Quality of Life". *The Repugnant Conclusion: Essays on Population Ethics.* Edited by Tännsjö, Torbjörn and Jesper Ryberg. Library of Ethics and Applied Philosophy: Vol. 15, Springer Netherlands, 2000, Dordrecht, pp. 7–22.

Rawls, John. *A Theory of Justice*. Belknap Press, 1971, Boston.

Sidgwick, Henry. *Methods of Ethics* (7th Edition). Palgrave Macmillan, 1962, London.

Singer, Peter. "Famine, affluence, and morality." *Philosophy and Public Affairs,* vol. 1, num. 3, 1972, pp. 229-243.

Sher, George (2014). Equality for Inegalitarians. Cambridge University Press.

Shulman, Charles, 2012, "Are pain and pleasure equally energy-efficient?" reflectivedisequilibrium.blogspot.com, http://reflectivedisequilibrium.blogspot.com/2012/03/are-pain-and-pleasure-equally-energy.html