

Is Objective Act Consequentialism Satisfiable?

Johan E. Gustafsson*

A compelling requirement on normative theories is that they should be satisfiable, that is, in every possible choice situation with a finite number of alternatives, there should be at least one performable act such that, if one were to perform that act, one would comply with the theory. In this article, I argue that, given some standard assumptions about free will and counterfactuals, Objective Act Consequentialism violates this requirement.

It is a view commonly held that, in any situation in which one might find oneself, one can always comply with morality. It's the basic idea behind

The Principle of Satisfiability

In every possible choice situation where the number of performable acts is finite, there is at least one performable act such that, if it were performed, it would not be wrong.¹

Why should moral theories conform to this principle? One answer is that normative theories should ideally guide one's actions and one can't be guided by a theory if one cannot comply with it, or more precisely, there is nothing one can do such that, if one were to do it, one would comply with the theory (Bykvist, 2007, 117). In this article, I shall argue that, given some standard assumptions about free will and counterfactuals, Objective Act Consequentialism violates the Principle of Satisfiability.

Before we go on, however, it might help to clarify the relation between the Principle of Satisfiability and moral dilemmas. *Moral dilemmas* are typically taken to be choice situations where each performable act would be

* I would be grateful for any thoughts or comments on this paper, which can be sent to me at johan.eric.gustafsson@gmail.com.

¹ Bykvist (2007, 116) puts forward a stronger proposal, more similar to the Principle of Strong Satisfiability (see the end of this article). Jackson and Pargetter (1986, 242) claim that 'surely no ethical theory should make it impossible for someone to do everything he or she ought.' See also Österberg 1988, 127 and Arrhenius 2000, 193. I have restricted the Principle of Satisfiability to situations where the number of performable acts is finite, because, in situations with infinitely many performable acts, it might be that the outcome of each performable act is worse than the outcome of some other performable act and then it's arguably plausible that every performable act would be wrong if it were performed; see Savage 1954, 18.

wrong if it were performed (Vallentyne 1989, 301, Sinnott-Armstrong 1996, 49–50, and Zimmerman 1996, 207–8). Assuming a finite number of performable acts, the Principle of Satisfiability rules out moral dilemmas in this sense. These dilemmas should be distinguished from prohibition and obligation dilemmas. *Prohibition dilemmas* are choice situations where every performable act is wrong; *obligation dilemmas* are choice situations where each of two incompatible, performable acts is obligatory (Vallentyne, 1989, 302; 1992, 118). Assuming a finite number of performable acts, these kinds of dilemmas are ruled out by the combination of the Principle of Satisfiability and

The Principle of Strong Normative Invariance

If an act is performable in a choice situation, then its normative status does not depend on which of the performable acts in the situation are performed.²

On its own, however, the Principle of Satisfiability does not rule out prohibition or obligation dilemmas. If the Principle of Strong Normative Invariance doesn't hold, there might be prohibition dilemmas where all performable acts are wrong but at least one of these acts wouldn't be wrong if it were performed; and, similarly, there might be obligation dilemmas where there are two incompatible, performable acts each of which is obligatory but at least one of which wouldn't be obligatory if the other act were performed. Hence there might be prohibition and obligation dilemmas that are not moral dilemmas.³ Nevertheless, prohibition and obligation dilemmas, as well as violations of the Principle of Strong Normative Invariance, need not be very worrying as long as the Principle of Satisfiability holds, because, if that principle holds, one might still be able to comply with morality.

According to

² Carlson 1995, 101. Prichard (1932, 26) proposes a similar idea but is mainly concerned with whether being obligatory is a property of the act (and thus a property that depends on the existence of the act).

³ If normative variance is possible, then there might be prohibition and obligation dilemmas that are not moral dilemmas and Vallentyne's taxonomy of moral dilemmas needs to be expanded with a further kind of moral dilemma: Let a *variance dilemma* be a choice situation where (i) there is at least one performable act that is not wrong, (ii) there are not two or more incompatible, performable acts each of which is obligatory, and (iii) each performable act would be wrong if it were performed.

Act Consequentialism

An act x is right in a choice situation if and only if the outcome of x is not worse than the outcome of any alternative act that is performable in the situation. And an act is wrong if and only if it is not right.⁴

In addition, according to *Objective Act Consequentialism*, the morally relevant outcome of an act is, roughly, what would happen if the act were performed.⁵ This addition is motivated by the basic tenet of Consequentialism: that morality is about making the world as good as possible (Parfit 1984, 24 and Carlson 1995, 56). This rough account of outcomes can be spelled out in a number of ways—for example:

The Principle of World Outcomes

The outcome of an act is the possible world that would be actual if the act were performed.⁶

The Principle of Future-State Outcomes

The outcome of an act is the total future state of the world which would obtain if the act were performed.⁷

For our discussion, the crucial feature of these principles is that they both take the outcome of an act to be something that would be actual (or the case) if the act were performed. Regarding subjunctive conditionals of this kind, we shall adopt the standard view that, if there is a closest (accessible) possible world where an act is performed and something is the case in that world, then it would be the case if the act were performed (Stalnaker 1968, 102 and Lewis 1973, 16). Moreover, we shall assume that one is sometimes able to act otherwise than one in fact does, and that, if one would have acted otherwise, some features about the choice situation up until one made one's choice would have been different. Perhaps some true historical proposition would have been false or some law of nature would have been broken. This follows from, for example, *soft determinism*: the claim that determinism is true but agents are sometimes able to act

⁴ Moore 1912, 31. Following Moore (1912, 12–18), we require that an act has to be within the agent's present voluntary control to be a performable alternative.

⁵ In Jackson and Pargetter's (1986, 233) terms, we assume actualism rather than possibilism. The argument of this article should also apply, changing what needs to be changed, to possibilist versions of Objective Act Consequentialism.

⁶ Carlson 1995, 10. For similar accounts, see Moore 1903, 147; 1912, 181; 1942, 558, Feldman 1986, 36, Sosa 1993, 101, Tännsjö 1998, 31, and Portmore 2011, 34.

⁷ Carlson 1995, 10. Compare Bergström 1966, 123–25. Carlson (1995, 56–57) argues that any way of taking the morally relevant outcome of an act to be the causal consequences of the act must, in order to be normatively plausible from the point of view of Objective Act Consequentialism, be axiologically equivalent to either the Principle of World Outcomes or the Principle of Future-State Outcomes.

otherwise. For instance, it might be that, if one had chosen otherwise, then, just before one made one's choice, one's credences or desires would have been slightly different.

Now, suppose that you are choosing between becoming an investment banker and becoming a voluntary worker. Any future choices after this career choice are beyond your present voluntary control. As a matter of fact, you will choose to become an investment banker, and this will make you rich. But, because you are selfish, you won't use your riches to do much good in the world. You would do more good if you were to become a voluntary worker instead. Accordingly, Objective Act Consequentialism yields that becoming an investment banker is wrong, and, since you will become one, it yields that, if you were to become an investment banker, it would be wrong.⁸ Suppose that the reason you will choose to become an investment banker is that you are selfish.⁹ And, in the closest possible world where you instead choose to become a voluntary worker, you are less selfish and more altruistic.¹⁰ Suppose moreover that, from the point of view of that world—that is, the closest possible world where you become a voluntary worker—the closest possible world where you become an investment banker isn't the actual world but a world where you are still altruistic.¹¹ And, in that world where you are altruistic and become an investment banker, you will do much more good with your riches than you could do as a voluntary worker. So, if you were to become a voluntary worker, it would be better to become an investment banker. Accordingly, Objective Act Consequentialism yields that, if you were to become a

⁸ I assume the standard view that subjunctive conditionals with true components are true; see Lewis 1973, 26–29, but compare Bennett 1974, 387–88; 2003, 239–40.

⁹ One might wonder whether this conflicts with your being able to become a voluntary worker. But, if we were to rule out that one sometimes can (in the same sense as in the doctrine that *ought* implies *can*) do things which one is in fact unmotivated to do, Objective Act Consequentialism would become too permissible. It's implausible that egoists can't sometimes act wrongly by failing to do some self-denying act they are selfishly motivated to avoid. See Parfit 1984, 14–15.

¹⁰ It might seem strange that someone could instantly change their motivation merely by making a choice. The idea, however, is only that you would have been more altruistic if you had made this choice. Your choice wouldn't cause your altruism; rather, the direction of causation would run in the opposite direction. See Lewis 1981, 116–17.

¹¹ It's not essential for the example that you would have had different desires if you had chosen to become a voluntary worker. Suppose instead that, in the closest world where you become a voluntary worker, you find it more credible that selfish people are happier if they choose to become voluntary workers, as opposed to investment bankers. That is, you would (contrary to your actual beliefs) believe that even selfish people enjoy helping others. Then, we suppose that, from the point of view of the closest world where you become a voluntary worker, the closest world where you become an investment banker isn't the actual world but a world where you still believe that selfish people enjoy helping others and also that you would selfishly enjoy effectively helping others by earning to give.

voluntary worker, it would be wrong.¹² Hence each performable act in the situation (becoming an investment banker, becoming a voluntary worker) would be wrong if it were performed.¹³

Consider a more general example with the same structure. Suppose that you face a situation in which you can either perform act *a* or perform act *b*. World w_1^* is the actual world, in which you perform *a*; world w_2 is the possible world where you perform *b* which is closest to w_1^* ; and world w_3 is the possible world where you perform *a* which is closest to w_2 .¹⁴ Furthermore, the outcome of *a* in w_1^* is worse than the outcome of *b* in w_2 , and the outcome of *b* in w_2 is worse than the outcome of *a* in w_3 . Given that you perform *a*, there are no further things under your voluntary control in the situation which would affect any of the aspects in which w_1^* and w_3 differ. The example can be illustrated by the following table, where the worlds' uneven horizontal distribution reflects their relative closeness:

| World | w_1^* | w_2 | w_3 |
|--------|----------|----------|----------|
| Choice | <i>a</i> | <i>b</i> | <i>a</i> |
| Value | 1 | 2 | 3 |

The value of the outcome of *a* is 1 and the value of the outcome of *b* is 2. Similarly, we have that, if you were to perform *b*, the value of the outcome of *b* would still be 2 but the value of the outcome of *a* would be 3. In this situation, performing *a* is wrong since it has a worse outcome than *b*. But, if *b* were performed, *b* would be wrong since it would have a worse outcome than *a*. So each performable act in this situation would be wrong if it were performed. Objective Act Consequentialism thus violates the Principle of Satisfiability.¹⁵

¹² Here, we assume *the Principle of Performability Invariance*: whether an act is performable in a choice situation does not depend on which of the performable acts in the situation are performed. We shall consider lifting this assumption later (see note 20).

¹³ This example doesn't trade on the much discussed problem that the outcome of an alternative act might depend on what other alternative acts are *also* performed in the same situation; see Goldman 1978, 188–89. That problem, unlike this one, can be blocked by the requirement that the relevant alternatives should be incompatible in pairs; see Bergström 1966, 33–36 and Gustafsson 2014, 586–87.

¹⁴ In the actual choice situation, you are in w_1^* already (where you will perform *a*), even though you can (but won't) perform an act such that, if you were to perform it, you would be (and would have been) in w_2 .

¹⁵ Moreover, in each situation we have considered, Objective Act Consequentialism yields that there is an act that is right which wouldn't be right if it were performed. Objective Act Consequentialism therefore violates the Principle of Strong Normative Invariance. These situations are, according to Objective Act Consequentialism, neither prohibition nor obligation dilemmas; they are variance dilemmas—see note 3.

One might object that the outcome of a can't be better in w_3 than in w_1^* , since this act is the same in both of these worlds. A central idea in Objective Act Consequentialism, however, is that the outcome of an act depends not only on the intrinsic properties of the act but also on how the act would interact with the rest of the world if it were performed. So, due to some difference between w_1^* and w_3 which is extrinsic to a , this act has different outcomes in these worlds. And the difference between these outcomes is such that it makes one of them better than the other.

One might also question the assumptions about the relative closeness of the possible worlds in my example.

First, in w_1^* there is a unique closest world where b is performed, and in w_2 there is a unique closest world where a is performed. This follows from *counterfactual determinism*: the claim that, for each performable act in a situation, there is a unique closest possible world where that act is performed.¹⁶ Without counterfactual determinism, some acts would lack an outcome if we were to accept the Principle of World Outcomes. We would then have to fall back on a more complicated account of outcomes, for example, an account where the outcome of an act is a probability distribution over states of affairs, the probabilities these states would have of obtaining if the act were performed. Then, with that probabilistic account of outcomes, one could restate my example without the assumption of counterfactual determinism.

Second, my example assumes that, if a is performed in the actual world and w_2 is the closest world where b is performed, then the actual world needn't be the world where a is performed which is closest to w_2 . Yet this assumption is fairly minimal given, for example, soft determinism. On soft determinism, if one were to act otherwise, either some law of nature or some historical fact would be false. Then, on any plausible measure of the closeness of possible worlds, we have that the worlds where a is performed must have at least some degree of remoteness from w_2 . And, if so, the worlds where a is performed might plausibly differ in how much their historical facts and laws of nature depart from those in w_2 . And then we can plausibly suppose that w_1^* and w_3 differ in some of the respects that make possible worlds more or less remote to w_2 so that w_2 is closer to w_3 than to w_1^* .¹⁷ One way to resist this upshot would be to rely on a closeness measure that favoured worlds that are exactly like the actual world up until

¹⁶ See, for example, Goldman 1976, 452, Carlson 1995, 12, Tännsjö 1998, 41–43, and Bykvist 2002, 46, 66n2. The assumption of counterfactual determinism also shows that my example doesn't trade on the problem with underspecified counterfactuals discussed by Vessel (2003).

¹⁷ Note also that, while w_1^* can be closer to w_2 than to w_3 , w_1^* could also be closer to w_3 than to w_2 . Johansson (2010, 297) presents a case with the latter structure for the extrinsic personal value of states of affairs.

the agent's choice except for an instant minimal local miracle enabling the different choice. Since everything else is fixed, a world where you choose one thing and the world closest to it where you choose otherwise would then only differ, up to the time of the choice, in what choice is made. So, from the point of view of the closest world where *b* is performed, the closest world where *a* is performed would be the actual world. Hence my example would be blocked. The trouble with this solution is that, if we keep your psychology fixed apart from your choosing otherwise, you wouldn't be in a coherent psychological state if you were to choose *b*. To see this, assume that your psychology is coherent in the actual world and that, given your credences and desires, you judge that you have most reason to perform *a*. But then, if we favour holding your psychology fixed when we measure closeness, the closest world where you perform *b* would be a world where your credences and desires still support the judgement that you have most reason to perform *a*, rather than *b*. It's implausible to assess the option of performing *b* by looking at a world where you choose *b* against your own judgement and in conflict with your credences and desires. On the other hand, closeness measures that likewise favours keeping other things fixed up until the agent's choice but which more strongly favour the agent's being psychological coherent wouldn't block my example. They wouldn't do so, because the minimal psychological differences between the actual world and the closest world w_2 where you coherently perform *b* could plausibly be greater than the minimal psychological differences between w_2 and a world where you coherently perform *a*.

One might also object that, if acts were individuated finely enough, there couldn't be a situation where the same act is performed with different outcomes in two possible worlds, which *a* is in my example.¹⁸ So, given a sufficiently fine-grained individuation of acts, my example would be impossible. With a more fine-grained individuation, we could replace *a* by two more specific acts: *a'*—which is performed in w_1^* but not in w_3 —and *a''*—which is performed in w_3 but not in w_1^* . Then *a''* would be an act such that it wouldn't be wrong if it were performed. But, given this individuation, only *a'* and *b* are performable in the situation in w_1^* and only *a''* and *b* are performable in the situation in w_2 .¹⁹ We then

¹⁸ One could, for example, adopt Davidson's (1969, 231) view that acts are events and 'events are identical if and only if they have exactly the same causes and effects.'

¹⁹ Wlodek Rabinowicz proposes the following example, reported in Carlson 1995, 100n30, where the performability of an act depends on what is done, given a conditional analysis of ability:

Suppose that Brown tries to lift a 100 pound weight, and succeeds. Then his action of lifting a 100 pound weight (at that particular time) is performable. But the nearest possible world where he does not lift this weight

have that a' , which will be performed, is wrong since it has a worse outcome than b and, if b were performed, b would be wrong since it would have a worse outcome than a'' . Hence, in the situation you face in the actual world w_1^* , we still have that there is no performable act such that it wouldn't be wrong if it were performed.²⁰

Finally, one might object that the Principle of Satisfiability is compatible with Objective Act consequentialism if it is amended as follows:

The Amended Principle of Satisfiability

In every possible choice situation where the number of performable acts is finite, there is at least one performable act such that, if it were performed *in that choice situation*, it would not be wrong.

If we individuate choice situations by, among other things, the outcomes of the performable acts, then Objective Act Consequentialism does not violate this amended principle. Consider your choice between a and b . In the actual world, where you perform a , the outcome of a is worse than that of b . But, in the closest world where you perform b , the outcome of your doing so is worse than that of a . So, in that world, the choice situation in which you perform b isn't the same as the one you face in the actual world, because the outcomes of the performable acts are different. Hence my example does not challenge the compatibility between Objective Act Consequentialism and the amended principle.

But this amended principle is of limited interest. It fails to capture the basic idea behind the Principle of Satisfiability: that, in any choice situation, there should be some performable act such that, if one performed it,

may well be a world where he wants and tries to lift it, but fails. A typical conditional analysis would then say that this action was not performable.

This example, however, is only an example of the performability of an act depending on whether one succeeds in doing what one chooses to do. It's not an example of the performability of an act in a situation depending on what one chooses to do in that situation. Bykvist (2002, 62) proposes an example where whether an act is an alternative in a situation depends on what act is performed in that situation. But, in his example, what one does in the situation doesn't influence the *performability* of any act in that situation; it only influences what acts count as *alternatives* in the situation. That is, it doesn't violate the Principle of Performability Invariance (see note 12)—it violates *the Principle of Alternative Invariance*: whether an act is one of the (relevant) alternatives in a choice situation does not depend on which of the performable acts in the situation are performed; see Gustafsson 2014, 590. Plausibly, whether an act is an alternative in a situation supervenes on what acts are performable in that situation.

²⁰ If we allow that the performability of acts in a choice situation might depend on what acts are performed in the situation, we could have cases where Objective Act Consequentialism and the Principle of Satisfiability conflict without any counterfactual instability in the outcomes of acts. Consider a situation where only acts a and b are performable; b has a better outcome than a ; but, if b were performed, a further act c would be performable and the outcome of c would be better than each of the outcomes of a and b .

one would comply with morality. Even if Objective Act Consequentialism satisfies the amended principle, you are still unable to comply with the theory in my example. While the amended principle ensures that there will be an act such that it wouldn't be wrong if it were performed in the situation you face, the amended principle doesn't ensure that you would comply with Objective Act Consequentialism if you were to perform that act, because, if you were to do so, you needn't have been in the same choice situation as the one you face in the actual world. If you were to perform *b*, then *b* would be wrong in the choice situation in which it would be performed. Hence the amended principle fails to capture the basic idea behind the Principle of Satisfiability. To capture that idea more fully, however, we need a stronger principle than the Principle of Satisfiability in order to take into account that one fails to comply with morality if one performs an act that is wrong jointly with an act that isn't wrong:

The Principle of Strong Satisfiability

In every possible choice situation where the number of performable acts is finite, there is at least one performable act *x* such that, if *x* were performed, then, in the situation in which *x* would be performed, no performed act would be wrong.

The Principle of Strong Satisfiability is stronger than the Principle of Satisfiability. Therefore, since Objective Act Consequentialism violates the latter, it violates the former too.²¹

References

- Arrhenius, Gustaf (2000) *Future Generations: A Challenge for Moral Theory*, Ph.D. thesis, Uppsala University.
- Bennett, Jonathan (1974) 'Counterfactuals and Possible Worlds', *Canadian Journal of Philosophy* 4 (2): 381–402.
- (2003) *A Philosophical Guide to Conditionals*, Oxford: Clarendon Press.
- Bergström, Lars (1966) *The Alternatives and Consequences of Actions: An Essay on Certain Fundamental Notions in Teleological Ethics*, Stockholm: Almqvist & Wiksell.
- Bykvist, Krister (2002) 'Alternative Actions and the Spirit of Consequentialism', *Philosophical Studies* 107 (1): 45–68.

²¹ I wish to thank Campbell Brown, Krister Bykvist, Erik Carlson, Richard Chappell, Barry Lee, Daniel Morgan, Paul Noordhof, Martin Peterson, Christian Piller, Douglas W. Portmore, Wlodek Rabinowicz, Tom Stoneham, Helen Yetter-Chappell, the audiences at the Practical-Philosophy-Group Seminar, University of York, 8 February 2016, and ISUS XIV 2016, Lille Catholic University, 6 July 2016, and an anonymous referee for *Analysis* for valuable comments.

- (2007) ‘Violations of Normative Invariance: Some Thoughts on Shifty Oughts’, *Theoria* 73 (2): 98–120.
- Carlson, Erik (1995) *Consequentialism Reconsidered*, Dordrecht: Kluwer.
- Davidson, Donald (1969) ‘The Individuation of Events’, in Nicholas Rescher, ed., *Essays in Honor of Carl G. Hempel: A Tribute on the Occasion of His Sixty-Fifth Birthday*, pp. 216–234, Dordrecht: Reidel.
- Feldman, Fred (1986) *Doing the Best We Can: An Essay in Informal Deontic Logic*, Dordrecht: Reidel.
- Goldman, Holly S. (1976) ‘Dated Rightness and Moral Imperfection’, *The Philosophical Review* 85 (4): 449–487.
- (1978) ‘Doing the Best One Can’, in Alvin I. Goldman and Jaegwon Kim, eds., *Values and Morals: Essays in Honor of William Frankena, Charles Stevenson, and Richard Brandt*, pp. 185–214, Dordrecht: Reidel.
- Gustafsson, Johan E. (2014) ‘Combinative Consequentialism and the Problem of Act Versions’, *Philosophical Studies* 167 (3): 585–596.
- Jackson, Frank and Robert Pargetter (1986) ‘Oughts, Options, and Actualism’, *The Philosophical Review* 95 (2): 233–255.
- Johansson, Jens (2010) ‘Being and Betterness’, *Utilitas* 22 (3): 285–302.
- Lewis, David (1973) *Counterfactuals*, Oxford: Blackwell.
- (1981) ‘Are We Free to Break the Laws?’, *Theoria* 47 (3): 113–121.
- Moore, G. E. (1903) *Principia Ethica*, Cambridge: Cambridge University Press.
- (1912) *Ethics*, London: Williams & Norgate.
- (1942) ‘A Reply to My Critics’, in Paul Arthur Schilpp, ed., *The Philosophy of G. E. Moore*, pp. 535–677, Evanston and Chicago: Northwestern University Press.
- Parfit, Derek (1984) *Reasons and Persons*, Oxford: Clarendon Press.
- Portmore, Douglas W. (2011) *Commonsense Consequentialism: Wherein Morality Meets Rationality*, New York: Oxford University Press.
- Prichard, H. A. (1932) *Duty and Ignorance of Fact*, London: Humphrey Milford.
- Savage, Leonard J. (1954) *The Foundations of Statistics*, New York: Wiley.
- Sinnott-Armstrong, Walter (1996) ‘Moral Dilemmas and Rights’, in H. E. Mason, ed., *Moral Dilemmas and Moral Theory*, pp. 48–65, New York: Oxford University Press.
- Sosa, David (1993) ‘Consequences of Consequentialism’, *Mind* 102 (405): 101–122.
- Stalnaker, Robert C. (1968) ‘A Theory of Conditionals’, in Nicholas Rescher, ed., *Studies in Logical Theory*, pp. 98–112, Oxford: Blackwell.
- Tännsjö, Torbjörn (1998) *Hedonistic Utilitarianism*, Edinburgh: Edinburgh University Press.
- Vallentyne, Peter (1989) ‘Two Types of Moral Dilemmas’, *Erkenntnis* 30

- (3): 301–318.
- (1992) ‘Moral Dilemmas and Comparative Conceptions of Morality’, *The Southern Journal of Philosophy* 30 (1): 117–124.
- Vessel, Jean-Paul (2003) ‘Counterfactuals for Consequentialists’, *Philosophical Studies* 112 (2): 103–125.
- Zimmerman, Michael J. (1996) *The Concept of Moral Obligation*, Cambridge: Cambridge University Press.
- Österberg, Jan (1988) *Self and Others: A Study of Ethical Egoism*, Dordrecht: Kluwer.