



The incentives account of feasibility

Zofia Stemplowska¹

Accepted: 8 February 2020 / Published online: 28 September 2020
© The Author(s) 2020

Abstract In *Utopophobia* Estlund offers a prominent version of a conditional account of feasibility. I think the account is too permissive. I defend an alternative incentives account of feasibility (of action). The incentives account preserves the spirit of the conditional account but qualifies fewer actions as feasible. Simplified, the account holds that an action is feasible if there is an incentive such that, given the incentive, the agent is likely to perform the action successfully. If we accept that ought implies feasible, then we should reject some normative requirements on agents that Estlund would accept in light of his more permissive conditional account. But we can still recognise normative requirements on individual and collective agents that, if complied with, would result in a world that is radically better than our own.

Keywords Feasibility · Utopophobia · Estlund · Lawford-Smith

Some instances of utopophobia may be due to an overly narrow understanding of when actions that can be required of individuals and groups are feasible. Such instances can then presumably be cured with more permissive accounts, that is accounts that would qualify more actions as feasible. In *Utopophobia* (2019) and his earlier work (2008; 2011), Estlund offers a prominent version of such a permissive account: the conditional account of feasibility. I think the account is too permissive. I will instead defend an alternative incentives account of feasibility (of action). The incentives account

✉ Zofia Stemplowska
zofia.stemplowska@worc.ox.ac.uk

¹ University of Oxford, Oxford, UK

preserves the spirit of the conditional account but qualifies fewer actions as feasible.¹ Simplified, the account holds that an action is feasible if there is an incentive such that, given the incentive, the agent is likely to perform the action successfully.

Since I will also assume (without argument) that ought implies feasible, it will follow that I would reject some normative requirements on agents that Estlund would accept in light of his more permissive conditional account. Though, as my account preserves the spirit of the conditional account, it is still an account that is friendly to accepting normative requirements on individual and collective agents that, if complied with, would result in a world that is radically better than our own.

Section 1 contains definitions. Section 2 outlines the conditional account. Section 3 argues that the conditional account, including Estlund's, overidentifies cases of feasible *individual* action. Section 4 offers the incentives account of feasibility as the solution. Section 5 argues that the conditional account, as developed by Estlund, overidentifies cases of feasible *collective* action. Section 6 offers the incentives account as the solution and defends it further in light of Lawford-Smith's analysis of the feasibility of collective action. Section 7 concludes. *Utopophobia* is a monumental achievement. It might be petty to focus on its imperfect account of individual and collective feasibility when this account is not central to Estlund's analysis. But, of course, all imperfections matter for utophiles and, in any case, Estlund is to blame for having written a book that has been so successful at persuading me about most of its core claims.

1 Definitions

I will talk of feasibility of individual and collective action but when it comes to individual action it is more intuitive to talk of ability to act. I will assume that ability analyses of individual action are feasibility analyses too. I define the key terms as follows:

- [a.] 'Action Φ is feasible' means 'Action Φ is feasible for agent X '.
- [b.] 'Action Φ ' means either 'action Φ done by an individual' or 'collective action Φ constituted by $\Phi_1 \dots \Phi_n$ done by $X_1 \dots X_n$ '.
- [c.] 'Agents $X_1 \dots X_n$ do $\Phi_1 \dots \Phi_n$ ', means 'each agent $X_1 \dots X_n$ respectively does $\Phi_1 \dots \Phi_n$ '.
- [d.] 'Agent X ' means either 'an individual agent' or 'a collective agent composed of all those individuals $X_1 \dots X_n$ whose each respective doing of $\Phi_1 \dots \Phi_n$ constitutes the collective action Φ '.
- [e.] 'Agent X does Φ ' means 'in context C at time T , agent X succeeds in performing action Φ thereby bringing about state of affairs S '.

¹ I draw here on my earlier work (2016) including, verbatim, when it comes to definitions offered in Sect. 1 as well as some of the phrasings of the incentives account. However, I also significantly revise the account I defended there and offer the revised version here in order to scrutinise Estlund's account of ability. At least one of the revisions I now adopt was urged for the earlier article already by the referee, George Rudebusch, though I was not ready at that point to concede he was right. I thank him more fully now.

- [f.] ‘Trying to do Φ ’ means ‘engaging in an appropriately sustained attempt to do Φ so that any failure to do Φ is not due to the agent giving up too early’ (cf. David Wiens 2016).
- [g.] In the case of collective action ‘an incentive I ’ means ‘a set of incentives such that there is an incentive $I_1 \dots I_n$ for each agent whose actions constitute the collective action’.

2 The conditional account

I agree with those who, like Erman and Möller (2019) as well as Southwood (2018), show that questions of feasibility may be posed to elicit distinctive answers to distinctive questions with distinctive standards of evaluation of the success of those answers. Here I am focusing on what Erman and Möller (2019) have called the functional role of feasibility.² In its functional role, as Southwood (Unpublished manuscript) explains, it divides actions into those that it makes sense for an agent to deliberate about performing and those that it does not.³

When is an action feasible for an agent? Clearly, it cannot be merely a matter of whether it is possible since the concept of possibility is already around and available instead. In any case, equating feasibility with possibility does not chime with the intuition that judgements of feasibility are judgements of what agents can accomplish in the circumstances they find themselves in rather than what freaky occurrences may happen to them. As Southwood and Brennan (2007: 8–10) have argued, for example, it may be possible that, through a fluke, a philosopher performs open heart surgery successfully, accidentally grafting a healthy artery to the coronary artery. Still, the action of performing successful open-heart surgery is not feasible for her. This pushes us towards associating feasibility simply with likelihood but this temptation too must be resisted. As Estlund (2008: 13–14; 2019: 27) has pointed out, it may be exceedingly unlikely that he would perform the chicken dance in front of his class, but the action seems feasible for him. I will take this intuition as a constraint on a successful account of feasibility of action.

The conditional account of feasibility deals with the chicken dance problem by making feasibility a function of possibility (Gilbert and Lawford-Smith 2012) or likelihood of success conditional on trying (Brennan and Southwood 2007; Estlund

² Brennan (2013), and Hamlin (2017) also problematise the many questions that feasibility could be about.

³ My reporting of Southwood’s views above is an oversimplification and he does not rely on the idea of whether deliberation ‘makes sense’ but whether it is ‘correct’. His paper is unpublished; the relevant quote reads: ‘The role of feasibility is to divide a subject’s set of potential doings into those that lie within the subject’s domain of deliberative jurisdiction and those that lie outside it, i.e. to divide the set of potential doings into those that it would be deliberatively correct (i.e. in accordance with constitutive norms of deliberative initiation) for the subject to deliberate about whether to do and those that it would be deliberatively incorrect (i.e. in violation of constitutive norms of deliberative initiation) for the subject to deliberate about whether to do.’ See also Erman and Möller (2019).

2011; Gheaus 2013: 450; Southwood 2016).⁴ The possibility version of the account still fails to deal with the open heart surgery case and I put it aside for that reason, but the likelihood version deals with this case well. According to the likelihood version: Action Φ is (more) feasible if agent X is (more) likely to Φ given that X tries to Φ . Estlund (2019: 94) himself offers an analysis of ability that falls into this category: Agent X is able to Φ iff were X to try (and persevere), X would tend to Φ .

The account deals with the chicken dance and the surgery cases because it discounts failures due to lack of trying but counts those due to some other interventions. As a matter of biographical fact, Estlund's chicken dance example and Southwood and Brennan's open-heart surgery case persuaded me to give up on accounts of feasibility that simply track likelihood (let alone possibility). But the conditional account is faulty too. It generates two problems. First, it sees too much feasibility in cases of individual motivational failure. Second, it sees too much feasibility in cases of collective action.⁵

3 Overcounting individual actions as feasible

A good account of feasibility, as I understand it here, should be able to distinguish between failures of action in case of mere unwillingness (the agent sufficiently controls her motivational state), like the chicken dance case, and cases where failures are due to agent's motivational inability (or, put differently, volitional incapacity). This is especially apparent if what we are after is an account of feasibility that helps sift actions into those that it does and does not make sense for the agent to deliberate about performing. But this is also important if we accept ought implies able/feasible.

What best illustrates the set of cases that represent motivational inability is itself controversial but a stylised indisputable (if scientifically fictional) case would be the failure to perform an action when performing it requires the firing of given synapses but the relevant synapses are, say, separated by a nail stuck in one's head. A more typical case is that of an agent not being able to stay awake once they have been awake for a sufficiently long time. Famous philosophical examples include Susan Wolf's (1990: 99) case of a woman confronted with an attacker who finds herself paralysed and unable to scream for help.

How does Estlund deal with the difficulty? He grants that there are cases of genuine inability such as what he calls 'clinical cases.' These are 'defined as motives that are commonly understood as chronic or temporary psychological disorders of the kind that call for medical care' (2019: 99). He also grants that there

⁴ See also Lawford-Smith (2012) and (again) Gilabert and Lawford-Smith (2012). But compare Gilabert (2017: 98), where he offers a qualification on the conditional account (this qualification makes an appearance in my discussion in Sect. 3 below).

⁵ There are other accounts of feasibility that I cannot review here. For example, to follow Southwood's (2018: 2–4) typology, in addition to the conditional accounts and possibility accounts there are also cost-based accounts (e.g. Rääkkä 1998: 33–40; Buchanan 2004; Miller 2013: 36–39) where an action is not feasible if the cost is too high.

may be other disabling motivations (e.g. love) but, plausibly, holds that many instances of action will not be subject to such disabling motivations. For example, as he suggests, voting one way or another or staying informed about policies is unlikely to be subject to them. When we are in the presence of ‘clinical cases’ the conditional account would offer us the wrong answer. Estlund’s strategy, having granted their existence, is to classify them as not analysable through the conditional account.

Estlund’s strategy should be seen as an ‘application-restricting solution’.⁶ Adopting the strategy is understandable if all Estlund wants to show is that, plausibly, there are cases that do not strike us, intuitively, as cases of motivational failure. But the move does not help us determine the shape or defend the conditional account of feasibility; the problem remains that the account is not up to the task of identifying correctly all cases of feasible action, since ‘trying’ may cover cases of both motivational failure and mere unwillingness. And since, as I argue in the next section, there is a better account available to us, we should reject the conditional account.

Before defending my alternative account, let me quickly gloss a couple of non-Estlund alternatives that I do not think work but that I cannot engage with here. Pablo Gilabert (2017: 97), for example, develops his conditional account of feasibility by adding a further clause to the ‘conditional on trying’ clause: ‘and A [the agent] can indeed try.’ The final clause is elaborated as the agent being able to decide to act. However, this leads to the question of what it means for an agent to be able to decide to act. Help, alas, won’t come from Kadri Vihvelin’s (2004: 443) attempt to draw a distinction between being able to act and being able to bring oneself to act. As Southwood and Gilabert (2016: 4) point out, there are problems with the two obvious readings of what ‘bringing oneself to act’ might involve: If bringing oneself to act requires deliberation (or deciding) to act then, implausibly, we must conclude that a person who acted on an impulse (e.g. caught a cricket ball thrown unexpectedly in her direction) was unable to bring herself to act. If, by contrast, bringing oneself to act stands for anything that is necessary to bring oneself to act, then it is no longer obvious that one can act without being able to bring oneself to act but, in any case, we have not advanced our analysis very far for our purposes of distinguishing volitional incapacity and mere unwillingness.⁷

The difficulty of separating cases of motivational inability and mere unwillingness arises, of course, not only in the context of conditional accounts of ability. David Wiens (2016), for example, who has rejected the conditional account, deals with the difficulty by suggesting that failure to act in the presence of a ‘good faith’ attempt is a case of motivational inability (while other cases would count as mere unwillingness). This is what we might call an internal account of motivational failure – internal because it requires us to look inside the agent to determine if their attempt is ‘good faith’. But if the standards for determining if it is are self-reporting

⁶ The term is due to Southwood and Gilabert (2016: 2 n.3).

⁷ Southwood and Gilabert (2016: 4–6) analyse the implausible implications of the second interpretation for the claim that an agent who is unable to bring herself to act may be able to act.

by the agent, then we face the problem of self-deception. There was a time when I thought that working with no daily time to relax was impossible for me; then I acquired children and realised I was wrong. If, on the other hand, the standards are external to the agent (e.g. how long she persevered), the worry is that we are left unsure if the failure is due to inability or unwillingness. My point is not simply that we may therefore not know whether we are in a given case in the presence or absence of a ‘good faith’ attempt but that the category of a ‘good faith’ attempt does not in fact help us distinguish between motivational inability and mere unwillingness.

4 The incentives account (1): individual action

Instead of the conditional account we should adopt the (existential) incentives account (see also Stemplowska 2016). Simplified, the account holds that an action is feasible if there could be an incentive such that it would incentivise the agent to try to perform the action and the agent is likely to succeed.⁸ The incentive may never be offered but it would, if offered under the circumstances in question, deliver the result. Still simplified (for now), we can say that:

- (1) action Φ is feasible if there is an incentive I such that, given I , X is likely to do Φ .⁹

What counts as an incentive that could exist in the relevant sense? I have in mind here something that if offered to the agent (as a reward or punishment) would motivate them even if, as it happens, it may not be possible to offer it. It must be the case that the incentive is conceptually possible but not that it is available to be offered. The incentive may exist, in other words, merely in the way that it is possible to imagine it coherently. If, say, a person would succeed in Φ -ing were they to be motivated by the thought of seeing a unicorn, then being able to glimpse a unicorn would qualify, as would the incentive to eliminate all suffering from the world or become loved by everyone.

It is also important that the incentive must not ‘act’ by directly altering any of the physical features of the person’s circumstance. This is to eliminate cases that influence the agents by affecting them directly – e.g. through smell, sound, vision or another sense since these, plausibly, may change rather than reveal agent’s abilities.¹⁰ For example, I have no doubt that, when hearing my child cry, I can stay awake for longer or run faster to reach them. Being able to stop the crying in such cases, however, is not merely a response to the incentive but also to the physical

⁸ I put aside how likely it needs to be.

⁹ The clause ‘is likely to do Φ ’ should be unpacked into ‘will try to do Φ and, given I , is likely to do Φ ’ such that the incentive does not merely make the trying likely but necessitates it. However, since an agent who won’t try will not be likely to do Φ as an agent, I use the condensed formulation here.

¹⁰ This generates a difficulty regarding what it means to offer an incentive without changing the person’s circumstances. Ideally, the offer should just appear in the person’s mind (to eliminate the need for accounting for the effect of communication of the incentive on the agent’s circumstances).

features of my environment. Being able to see or smell another person or an object may affect how a person acts but it amounts to a change in the features of agent's circumstances and, thus, does not reveal whether the agent has the ability to act in circumstances that do not have such features.

There is also a question of whether the credibility of the incentive must be assumed (e.g. whether we must assume that the agent finds it credible) or whether the fact that an agent does not believe in it (e.g. being offered a glimpse of a unicorn would not motivate me given my beliefs about unicorns) is itself an indication that this particular incentive does not exist. I am inclined to think that the first offers a better test of feasibility but, ultimately, settling it would make a difference only for a very narrow range of cases (so I put it aside): cases where an agent can be motivated only by impossibilities (e.g. glancing at a unicorn) or agents who find no incentives credible.

The incentives account delivers the right answer, at least for average agents, in the case of the chicken dance (if you give me £1m, I will dance), the open heart surgery (no matter what you give me, I am unlikely to succeed), the case of utter sleep deprivation and debilitating clinical phobia (no matter what is on offer, I cannot stay awake beyond some point). It would also, as can be expected, reveal some cases that would normally be classes as those of clinical phobias as cases not of inability but motivational failure (a homeless mother of a baby who has a clinical phobia of spiders may – depending on the severity of her case – be able to touch a spider if she were to receive housing for her and the baby as a result).

However, the incentives account seems to give us the wrong result in cases such as the unwilling murderer. There may be no incentive that would make someone murder someone else but this is reflective of their choice rather than motivational inability. Of course, we should expect such cases to be very rare. Many of us, I suspect, are capable of being incentivised into committing a whole range of acts, given the right incentives, since the incentive itself need not be self-regarding and, in fact, there is no limit on how feasible the incentive is – unicorns, saving the world, making the person who is about to be murdered happy, etc. would all qualify. But suppose some resist the action because they think it is the wrong thing to do (whether or not their standards of wrongness are accurate). If it is the case that, had they not seen it as wrong, there would be an incentive that would make the agent likely to succeed, then the action remains feasible for them.¹¹ This can be captured as follows:

- (2): Action Φ is feasible if there is an incentive I – or had the agent X not seen Φ as wrong there would be I – such that, given I , X is likely to Φ .

¹¹ There may be more difficult cases. Suppose had a person not seen Φ -ing as wrong there would be no incentive that would make her try – because she would be the type of person who is unresponsive to incentives – but, while she sees Φ -ing as wrong, it is only that she sees Φ -ing as wrong that explains why she does not Φ . As a matter of the psychology of humans, this is possible: altering one belief may alter others in various ways. But the counterfactual clause I propose is meant to capture the fact that seeing Φ -ing as wrong is the only cause of the agent not Φ -ing: the case is not overdetermined.

We should not be too surprised that the incentives account needs such a modification. The analysis of feasibility in terms of responsiveness to incentives is reminiscent of reason-sensitive accounts of moral responsibility.¹² Such accounts of moral responsibility, however, are partly attractive because they allow us to sidestep the question of when an agent is unable to act otherwise and when merely unwilling but endorsing the action she performs – either way she gets to be morally responsible. When it comes to feasibility of action, however, we need the proposed modification captured by (2). However, if choosing between the conditional account and the incentives account without the modification, I would still recommend the incentives account as more plausible, even if it would mean biting the bullet on the unfeasibility of performing an action one finds wrong no matter what incentive is on offer.¹³

Against Estlund's conditional account, the incentives account proves itself in the case of individual action. But since Estlund's focus in *Utopophobia* is, by design, on cases of individual inaction that we would intuitively classify as failures of motivation, adopting the incentives account does not affect Estlund's case that even the moral requirements that we know for sure won't be followed by agents who can't bring themselves to follow them (but are able to perform what the requirements demand) apply to those agents. There is more at stake in cases of the feasibility of collective action. Some cases of alleged utopophobia may simply rely on a better account of the feasibility of collective action than Estlund offers.

5 The incentives account (2): collective actions

Judgements of the feasibility of collective action are difficult when what one agent can achieve depends on what other agents do. An iconic example of one such case offered by Estlund (2019: 211–214) involves two agents:

Slice and Patch Go Golfing: 'Suppose that unless the patient is cut and stitched he will worsen and die (though not painfully). Surgery and stitching would save his life. If there is surgery without stitching, the death will be agonizing. Ought Slice to perform the surgery? This depends, of course, on whether Patch (or someone) will stitch up the wound.' Also, '[s]uppose that each doctor reasonably and truly believes that the other will not be there [at the hospital] to help the patient.' Slice and Patch are each going golfing.¹⁴

¹² With the most prominent one being that of Fisher and Ravizza (2000). But for a beautiful analysis of reason sensitivity see also Nomy Arpaly (2004).

¹³ I think the unmodified version, so to speak, is now endorsed by Southwood (Unpublished manuscript). An earlier mention of robustness analysis as a possible route to analyse conditionality, alongside the 'application-restricting solution', also appears in a footnote in Southwood and Gilbert (2016: 1 n.2).

¹⁴ This is a composite case on the basis of *Utopophobia*, pp. 211 and 214. I would like to put on record that, spookily, Estlund's earlier version of the example of golfing doctors preceded the following event: In 2015, the actor Harrison Ford crash landed an airplane on a golf course. The first person to reach Ford was a spinal surgeon who happened to be on the course, who could then do 'a brief exam' (Carroll and

Do Slice and Patch have the ability to save the patient? It seems safe to conclude that there is no collective *agency* here. Following Estlund (2019: 219), we can say that ‘a group is not an agent if its outputs are not determined in certain systematic ways by aggregating certain preferences, judgements, or choices of the group’s members.’ Although there is no collective agency, Estlund thinks that we are in the presence of ‘plural ability’ (and can make ‘plural requirements’ of the collective). For Estlund (2019: 247), a set of agents has plural ability to Φ iff that set would tend to succeed conditional on plural trying (e.g. conditional on them plural-setting out to Φ). What is plural setting out/plural trying? Estlund (2019: 215) does not offer a full analysis, but he does explain that ‘I will limit myself to examples in which the relevant agents know (or could easily come to know if they tried) what they need to know in order to perform the action in question. So, for example, I assume that Slice knows or could know whatever she would need to know in order to do the surgery, and Patch has the requisite knowledge to stitch’.

If Slice and Patch were to set out/plural try, would they succeed given their knowledge? To answer this question, consider a different case:

Buttons: There are two individuals who are strangers. Each is located in a separate room equipped with an identical row of 1000 non-sequentially numbered buttons. To save the life of a third party, each agent must press the button of the same number as the other agent. Each agent can press only one button. They cannot communicate with each other and do not know what number button the other agent intends to press.

Each individual in *Buttons* knows how to press the button. If they were both to press the button of the same number, they would succeed in saving the third party. Perhaps we should conclude that the individuals in *Buttons* know enough to set out together? But this seems entirely counterintuitive: the point of their condition is that they do not know enough to set out together. To make it a bit more concrete what they each do not know is which concrete action (or one of concrete actions) to perform given what concrete actions others will perform. Given their evidence, individual agents do not know how to contribute to the goal of saving the life so they cannot (individually or) together set out to do so.

Is *Slice and Patch* different? It does not seem so. While each doctor knows how to perform her action, recall that it is also the case that ‘each doctor reasonably and truly believes that the other will not be there to help the patient’. Thus, neither Slice nor Patch know, given their evidence, how to contribute to the goal of saving the life. True, if they *together* set out (e.g. if both went to the hospital), they would succeed but they cannot set out together given what each ‘reasonably and truly’ believes: given their evidence, each is reasonable unaware a course of action open to him/her that would contribute to delivering the outcome of saving a life.

We should conclude, then, that while the individuals from the relevant set of agents know how to perform some component of the desirable collective action,

Footnote 14 continued

Pengelly 2015). Clearly, *Stich and Patch* could be more imaginative about the role of the patient in obtaining healthcare.

they do not know enough to set out to together and try. If they were to set out together – each agent performing the action that together adds up to saving the life (the contributory Φ_1, \dots, Φ_n component parts of Φ) – they would tend to succeed. But the fact that they cannot set out to try is relevant to our assessment of whether they are able to perform the action – e.g. whether the action is feasible for them to perform.¹⁵

Estlund (2019: 247) disagrees. As he elaborates on plural ability, ‘[t]he question is not whether they are able, all together, to *set out* to do their parts. The setting out (or initial volitional process) is not something about which we are inquiring whether the agent is able to do it.’ The question is only whether they would tend to succeed if they were to set out together/to engage in plural trying. But this claim is left undefended. It may be that it is meant to draw support from the analogy with individual feasibility where, on the conditional account, the question is not whether an agent can try but whether she will likely succeed in performing a given action *if she tries*. But just as we should reject the conditional account in the case of individual ability, we should reject it in the case of collective ability. It is odd to think that agents may be plurally able when they cannot set out together. Unless plural ability is merely meant to denote the idea that all of them acting together would do the (required) trick, we should inquire into the ability of the agents to act together.

To see just how counterintuitive an account of plural ability that does not require agents being able to set out together is, consider another case:

Synchronic clapping. All able-bodied agents round the world need to clap their hands at exactly the same time tomorrow to produce a mighty sound.

Given the state of the world as it is, it strikes me as entirely unfeasible.¹⁶ Of course, if the agents were to try at the same time to clap, they would succeed in producing a mighty sound (at least in cities). But they cannot set out to try to clap tomorrow at the same time and the action is unfeasible.

This is not to say that Estlund’s account does not have an intuitive appeal. This is best illustrated if we think of his nuclear threat case:

¹⁵ This raises the question whether we should say that the individual action of *slicing in order to save the life in these circumstances* is one that Slice is able to perform. Clearly, she can slice and were she (and Patch) to arrive in hospital, she would perform the action of stitching that would save the life. It therefore appears deeply counterintuitive to say that she is not able to perform the action of *slicing in order to save the life in these circumstances*. But I think that we should bite the bullet and tidy up our intuitions. Recall that we are here trying to analyse individual ability/feasibility applied to *action* rather than mere behaviour or causal capacity. Given what she is reasonably aware of, she does not know how to be in a position to perform the action and so she is not able to do it intentionally. Just as the person locked in the cell is unable to escape even though she will easily escape if she presses a button hidden behind the third brick to the right of the window by the ceiling (that she cannot know about) or just as a person standing in front of Blackwell’s Bookshop in Oxford with no money is unable to buy a book by Bertrand Russell even if she will if she stumbles across £20 on the floor. I return to it in Sect. 6.

¹⁶ This unfeasibility is compounded by the further unfeasibility of having collective agency trying to coordinate them – even if they tried they would have failed, but this is not my focus here.

‘Mutual Assured Destruction: Whether the other party also does this or not, each of two hostile countries threaten massively deadly and wholly pointless destructive nuclear response to any nuclear attack from the other. It is not clear that either is morally required to withdraw this threat unilaterally if it seems necessary and effective in order to securely deter a first-strike attack. Suppose, in addition, that one or both countries reasonably, but mistakenly believes that the other country may attack either unprovoked or under a blameless misunderstanding, unless there is this deterrent.’ (Estlund 2019: 241).

With Estlund, I would like to (but cannot) say that this world is morally worse than a world in which such mutual destruction is not threatened. Given the intuitive pull of this conclusion, it is thus natural for us to think that humanity or at least both countries are collectively failing to follow moral requirements that they are able to follow. But if there is no account of the path that shows how individuals who make up the collectives could undertake to deliver the desired result, then the collective is not able to dismantle the threat since it is not able to set out to do so.

6 The incentives account of collective (and individual) feasibility

Where does this leave us? Agency is needed for feasibility in cases where what one person needs to do depends on what others are doing. If there is no agent, then there is no realistic account of what would be involved in having everyone clap hands together *at the same time*. To succeed in cases such as those, we need coordination, luck or magic. To succeed in different types of cases, where the task is simple and there is no adverse effect of doing the wrong thing for a given individual, we may simply need persistence, even blind persistence, from the individuals involved. But leaving luck, blind persistence, and magic aside, what coordination is meant to deliver is sufficient knowledge of what contributing action to undertake (and how to perform it). This knowledge may come from agents being told what to do or being able to analyse how their involvement can result in the desired outcome.

For example, if we return to cases of individual ability, agents need to know that a given action – pressing a third brick from the top of the fire place in the prisoner’s cell – will result in a given outcome – opening a secret passage out of the cell. If the prisoner has this knowledge, given other standard assumptions, it is feasible for her to escape; if she lacks it, it is not feasible.¹⁷ How to capture the fact that the

¹⁷ The final observation may be thought to reveal the superiority of the conditional account over the incentives account that was obscured above. On the conditional account, in line with our intuitions, it is not feasible for the prisoner to escape the cell if he does not know about the secret passage: were he to try, he would likely fail (since it would not occur to him to press the brick). Do we get the same result with the incentives account? We may worry (as I had in the past) that the incentive alone may carry information that renders it likely for the person to succeed, thus rendering action feasible in line with the account when it is not. For example, if we offer someone to whom it would otherwise never occur an incentive to press the third brick from the top, we thereby introduce the idea to do it, rendering it feasible. But the solution – as I now see – is to simply stipulate that the incentive that tests the feasibility of Φ -ing does not carry with it any information beyond what is available to the agent under the circumstances we consider. In the prisoner case then we must ask whether there is an incentive such that if offered the agent would

feasibility of collective action will depend on coordination or other ways of delivering to agents knowledge of which actions to perform? I suggest that we can capture it by considering whether agents are likely to try to perform actions that together, if performed, are likely to result in them succeeding. And since I wish to avoid tackling the question of whether any individual contribution to Φ is causally necessary, I simply specify that the agents trying to do their contributing parts ($\Phi_1 \dots \Phi_n$) make it likely that Φ results.

The incentives account of feasibility: Action Φ is (more) feasible for agents $X_1 \dots X_n$ iff there is an incentive I (or had the agents $X_1 \dots X_n$ not seen doing $\Phi_1 \dots \Phi_n$ as wrong there would be I), such that, given I ,

[1.] agents $X_1 \dots X_n$ are likely to do $\Phi_1 \dots \Phi_n$.

and

[2.] doing $\Phi_1 \dots \Phi_n$ is likely to result in Φ .

This proposal narrows the range of cases where we would identify collective feasibility, but does it narrow it far enough? In her important analysis of the feasibility of collective action, Lawford-Smith (2012) suggests that we should narrow the range of cases even further. The starting point of her analysis, plausibly, is that, as she puts it, '[m]embers [of collectives] have abilities, and these are aggregated to determine a collective ability' (463). To illustrate how her account poses a challenge to the incentives account, let me focus on her discussion of whether the German military had the ability to overthrow Hitler given that it was 'physically close to Hitler in a way that few other groups were... It had plenty of weapons, and strategic training' (463):

Military coup: 'Loyalty to Hitler was extremely fierce. The penalty for treason was severe—in all likelihood death. There were spies and informants everywhere. This means that no soldier could have (without high risk of death) started planning and strategizing in the way required to initiate a successful coup. If you don't know who you can trust, and the chances are that you can't trust many people, the risks of trusting anyone are too high. Furthermore, soldiers swore individual oaths of allegiance to Hitler himself, which were regarded as extremely serious. This would have made it difficult for individuals to even conceive of conspiring against Hitler.' (464)

Lawford-Smith (2012: 464) concludes that '...closer inspection of the claim that the German military could have overthrown Hitler reveals that it is probably false. The military had the ability to overthrow Hitler if the soldiers making up the military

Footnote 17 continued

succeed in escaping. Given standard assumptions about people's expectations and knowledge, there likely is not, since, even if the prisoner were offered more than freedom, it would not occur to her to press the third brick from the top. I am grateful to George Rudebusch for comments on a different text that helped me develop my views here.

each had the ability to do their parts in overthrowing Hitler. But they didn't: the conditions prevented it.'

Why does Lawford-Smith think that individuals lacked the ability (and thus the collective did)? 'If one tried to begin planning the coup, he would soon enough confide in an informant, and the price of that would be death. This is true for any soldier. Most people would say the soldiers didn't really have the option of planning a coup, even though it is true that there's something they could have—very recklessly—done.' (Lawford-Smith 2012: 464). To see the underlying thought of individual ability at work here, consider what she says about another case that is reminiscent of *Slice and Patch*.

Piano. Four individuals who constitute a piano removal company are present and all are necessary to pick up a piano that is weighing on a child. However, each of them reasonably believes that not all others will lift, rendering his lifting futile.¹⁸

Lawford-Smith (2012: 466) thinks that the individuals, given their reasonable beliefs, are not required to lift but, as with Estlund's diagnosis in *Slice and Patch*, she thinks that each individual has the ability to contribute his share of lifting and the group has the ability to lift the piano. So Lawford-Smith's analysis of individual ability/feasibility seems to align with Estlund's in that she does not require that the individual knows how to contribute, in light of the evidence available to her, to the collective goal in order for us to say that her contributor action is feasible for her. However, she also resists the thought that to establish collective ability we need to analyse what would happen if all individuals together set out to try to bring about the desired outcome (of course if they all did, they would likely overthrow Hitler).

The incentives account of feasibility delivers different results. It does so because it identifies differently individual ability (feasibility) and also how to aggregate it into collective ability (feasibility). Regarding individual ability, the results are most clear in the case of *Slice and Patch*. There is no incentive that could be offered to Slice (and Patch) such that, given the incentive, Slice would likely do the contributory action (of going to the hospital and slicing). This is because Slice (and Patch) do not know what to do given their evidence. Of course, in desperation, they may go to the hospital but such an action is not likely if there is an opportunity cost; if they happen to go to the hospital and thereby succeed in saving the life this is by fluke rather than because the action was feasible all along.¹⁹ The action, in *Piano*, of lifting the piano off the child is slightly more likely since, if you are standing next to the piano, you are likely to try to lift it if you can think of nothing else to do (and if trying to lift it when others do not isn't harmful). But if we change the case a bit (and Lawford-Smith's rigorous analysis provides an illuminating taxonomy of the full range of cases) such that trying to lift the piano when others do not would kill the child, the individuals are not individually able to contribute to the collective

¹⁸ Lawford-Smith (2012) offers various versions of Piano including one that is exactly like Slice and Patch in that agents cannot coordinate and trying to lift the piano when others do not will cause more harm than it alleviates.

¹⁹ See also fn. 15 above.

goal. Finally, in the case of the *Military coup*, the judgement is complicated since it relies on empirical assessment. But, putting aside who is right about the history of the second world war, I think that there will be instances such that Lawford-Smith's analysis will find individual disability when I see ability. Even when the risk of death is substantial for many agents there is an incentive (it may be glory, Hitler's love [weird though this would be], saving their families, etc.) such that they would likely try to coordinate their actions to overthrow Hitler. If they knew that the lives of their families, say, were at stake, they would likely take some steps in the right direction (trying to identify other people who are on their side; avoiding appearing fanatically committed to the cause, etc.). Although taking such steps would be extremely risky in circumstances full of informers (and there is a further question, therefore, whether undertaking them would be morally required) we know – given that a plan to overthrow Hitler was created and almost pulled off successfully – that engaging in actions that could contribute to the coup was possible. The high risk of death faced by the soldiers does not indicate their lack of ability. It explains why in the circumstances at hand they were (many of them in an unblameworthy manner) unmotivated to try to coordinate their action. If a different incentive structure would have made them take the risk, taking the risk was feasible for them.

Moreover, I think that we should analyse collective feasibility by asking, with Estlund, what would happen if all relevant agents took the relevant actions simultaneously (though, unlike for Estlund, it must be the case that it is possible for them to set out together to do so). If the relevant agents took these steps, they would likely – though this is speculative – succeed in overthrowing Hitler. Thus the group had the ability to overthrow him.

Lawford-Smith (2012: 464) entertains, and rejects, a possibility along these lines: 'We could say that a collective action is not ruled out if the parts of it are not ruled out for any member of the collective; and the parts of it are not ruled out for any member of the collective if the member has an action available to her that could produce her doing a part. That would mean only the soldiers' forcible prevention from doing their parts would suffice to genuine collective inability. But that seems much too strong.' I agree that this feels too strong if we elaborate an individual 'has an action available to her that could produce her doing a part' as one where even in *Buttons* individuals have the action of pressing the same button available to them. If, however, we circumscribe it to include only actions that, given a powerful-enough incentive, the agent will try and likely succeed, I do not think it is too strong: it reveals to us that we are in the situation we are because of a motivational failure rather than an inability on the part of the agent to perform the action.

In effect, we have three models of group ability to choose between. Estlund's model suggests that just as long a group would likely Φ if each agent who knows how to $\Phi_1 \dots \Phi_n$ did $\Phi_1 \dots \Phi_n$ (even if they did not know that $\Phi_1 \dots \Phi_n$ was needed from them to deliver Φ), the group has the ability. Against this, both my and Lawford-Smith's models hold that a further condition must be met. For Lawford-Smith the condition is that for each agent performing $\Phi_1 \dots \Phi_n$, the risks are not too high. For me, the condition is that there is an incentive that, given it, the agents are likely to try to $\Phi_1 \dots \Phi_n$.

7 Utopia

Where does this all leave us regarding normative requirements on collectives? Estlund's are in a sense the most ambitious. When we fail in cases like *Slice and Patch*, *Buttons*, and the more realistic ones like *Military Coup* and *Mutual Assured Destruction*, we violate the plural requirements that apply to us (even if it does not follow that the individuals can be individually blameworthy). Lawford-Smith's requirements are less ambitious since she rejects the possibility of group ability that is as divorced from individual ability to set out to try to do her part as Estlund suggests. That said, she also accepts the possibility that a group may be to blame when no individual is: in the *Piano* case, the group has the ability since each individual has the ability to do his share (without facing risk of death) but if they reasonably believe that not all will do their necessary share of lifting, they are morally off the hook.²⁰ With the exception of the *Military coup* and perhaps the *Piano* (depending on the variable details of the case), I think that collective action is not feasible in the scenarios. I think that there is no collective moral failure here (and no individual one either). If we are in suboptimal position as a result, rather than in utopia, in such cases this is because the world makes utopia unachievable for us, except through a fluke.

The results the incentives model delivers in cases such as *Buttons* and *Slice and Patch* rely on the absence of coordination, even communication between the parties.²¹ In the real world, of course, communication channels exist and coordination, even if initially absent, can often be achieved. Does this mean that in the real world all collective action is possible? Alas. The absence of coordination is sufficiently widespread to make it worthwhile to develop tools for assessing situations in which it is lacking. When can we say that humanity is failing to eradicate poverty and stop catastrophic climate change rather than being condemned to it? When are states failing to use their ability to collectively disarm their nuclear arsenals? I think that the incentives account of feasibility gives us the right answers here or, more precisely, direct us to ask the right questions of empirical sciences.²²

It might be hoped that, given enough time, all collective action could be brought about. Even synchronised clapping of all able-bodied adults may perhaps be achievable (even if not tomorrow). Should we, therefore, think that all is feasible or should we think that the incentives account of feasibility is wrong because, intuitively, many actions are not feasible but my account implies that they are? The account – alongside Estlund's conditional account of ability – ignores, the objection goes, the real source of infeasibility: lack of appropriate motivations. It suggests that real world collective actions are feasible when in fact people are not motivated to

²⁰ They are off the hook in that they fulfil their individual duties as these duties are sensitive to evidence. 'So if any member has good reason to believe that at least one of the others will fail to take a share, he will not be obliged to do his own share. The collective action can fail without any member being to blame.' (Lawford-Smith 2012: 465-6).

²¹ I am grateful for discussion to Kasim Khorasane, private correspondence.

²² But see Emily McTernan (2019).

perform them or are not coordinating (though they could). Shouldn't we want our judgements of feasibility to track that?

But the incentives account – just as to some extent the conditional account – can track the presence or absence of appropriate motivations.²³ If we consider what is feasible for the Labour party in the circumstances it finds itself, we will see that, say, the action of introducing a permissive immigration system is not feasible because the voters are opposed to it. But it does not follow that the action is not feasible for the UK: British citizens could coordinate and achieve such openness: it is enough that each person votes the right way.

Although I am unable to prove it, let alone show it here, my hunch is that feasible courses of action are open to collectives and individuals such that, if followed, our world would be close to perfectly just if not perfectly just. I think, with Estlund, that there is important value in understanding that morality may require us to build such a world and that, when, predictably, we fall short again and again and again, it does not shrink its aspirations for us. With Rawls (2001: 37–8), I find it consoling to think that although humans are demonstrably capable of terrifying evil, we are also capable and required to build a utopia even if we haven't.

Acknowledgements I am grateful for comments that greatly improved the paper to an anonymous referee, Ben Jackson, Henrik D Kugelberg and Nicholas Southwood. I am also grateful to the audience of the LSE workshop (May 2019) and to Simon Caney, Matthew Clayton, Alan Hamlin, Jonathan Quong and George Rudebusch for comments on an earlier version of this argument (see also footnote 1).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arpaly, N. (2004). *Unprincipled virtue: An inquiry into moral agency*. Oxford: Oxford University Press.
- Brennan, G. (2013). Feasibility in optimizing ethics. *Social Philosophy and Policy*, 30(1–2), 314–329.
- Brennan, G. & Southwood, N. (2007). Feasibility in action and attitude. In J. Josefsson and D. Egonsson (Eds.), *Hommage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz*. Department of Philosophy, Lund University
- Buchanan, A. (2004). *Justice, legitimacy and self-determination: moral foundations for international law*. Oxford: Oxford University Press.
- Carroll, R., & Pengelly, M. (2015). Harrison Ford 'saved several lives' by landing plane on golf course-witness. *The Guardian*. March 6. Available from: <https://www.theguardian.com/film/2015/mar/06/harrison-ford-plane-crash-leaves-actor-seriously-injured>
- Erman, E., & Möller, N. (2019). A world of possibilities: the place of feasibility in political theory. *Res Publica*, 26, 1–23.

²³ As can the conditional account. See for example Gilabert (2017).

- Estlund, D. (2008). *Democratic authority: A philosophical framework*. Princeton, NJ: Princeton University Press.
- Estlund, D. (2011). Human nature and the limits (if any) of political philosophy. *Philosophy & Public Affairs*, 39(3), 207–237.
- Estlund, D. (2019). *Utopophobia: On the limits (if any) of political philosophy*. Princeton, NJ: Princeton University Press.
- Fisher, J. M., & Ravizza, M. (2000). *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.
- Gheaus, A. (2013). The feasibility constraint on the concept of justice. *The Philosophical Quarterly*, 63(252), 445–464.
- Gilbert, P. (2017). Justice and feasibility in political utopias. In M. Weber & K. Vallier (Eds.), *Political utopias: Contemporary debates*. Oxford: Oxford University Press.
- Gilbert, P., & Lawford-Smith, H. (2012). Political feasibility: A conceptual exploration. *Political Studies*, 60(4), 809–825.
- Hamlin, A. (2017). Feasibility four ways. *Social Philosophy and Policy*, 34(1), 209–231.
- Lawford-Smith, H. (2012). The feasibility of collectives' actions. *Australasian Journal of Philosophy*, 90(3), 453–467.
- McTernan, E. (2019). Justice, feasibility, and social science as it is. *Ethical Theory and Moral Practice*, 22(1), 27–40.
- Miller, D. (2013). *Justice for earthlings: Essays in political philosophy*. Oxford: Oxford University Press.
- Räikkä, J. (1998). The feasibility condition in political theory. *Journal of Political Philosophy*, 6(1), 27–40.
- Rawls, J. (2001). *Justice as fairness*. Cambridge: Harvard University Press.
- Southwood, N. (2016). Does “ought” imply “feasible”? *Philosophy & Public Affairs*, 44(1), 7–45.
- Southwood, N. (2018). The feasibility issue. *Philosophy Compass*, 13(8), 1–13.
- Southwood, N., & Gilbert, P. (2016). Ability and volitional incapacity. *Journal of Ethics and Social Philosophy*, 10(3), 1–8.
- Southwood, N. (Unpublished manuscript). ‘Feasibility as Deliberative Jurisdiction’, talk given at Oxford, May 2019.
- Stemplowska, Z. (2016). Feasibility: Individual and collective. *Social Philosophy and Policy*, 33(1–2), 273–291.
- Vihvelin, K. (2004). Free will demystified: A dispositional account. *Philosophical Topics*, 32(1/2), 427–450.
- Wiens, D. (2016). Motivational limitations on the demands of justice. *European Journal of Political Theory*, 15(3), 333–352.
- Wolf, S. (1990). *Freedom within reason*. New York: Oxford University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.