The Non-identity Fallacy: Harm, Probability and Another Look at Parfit's Depletion Example

M. A. ROBERTS
The College of New Jersey

The non-identity problem is really a collection of problems having distinct logical features. For that reason, non-identity problems can be typed. This article focuses on just one type of non-identity problem, the 'can't-expect-better' problem, which includes Derek Parfit's depletion example and many others. The can't-expect-better problem uses an assessment about the low probability of any particular person's coming into existence to reason that an earlier wrong act does not harm that person. This article argues that that line of reasoning is unusually treacherous in that it makes not just one hard-to-detect error in what is done with the relevant probability assessments but rather alternates between two. We sort out one fallacy only to fall, against all odds (as it were), into a second. By avoiding both errors, we become able to discern harm in cases in which the can't-expect-better problem argues there is none. We will then be in a position to set aside the can't-expect-better problem as an objection against the person-based intuition that acts that are 'bad' must be 'bad for' at least some existing or future person.

I. THE NON-IDENTITY PROBLEM

The person-affecting, or person-based, intuition, as Parfit puts it, is the idea that 'what is bad must be bad for someone'.[1] Naturally and pre-analytically, we are sure to wonder how it is that a choice that is not bad for anyone – a choice that does not *harm* anyone who does or ever will exist – could nonetheless turn out to be morally wrong.

The person-based intuition seems compelling whether spelled out in consequentialist terms or otherwise. On the consequentialist side, it suggests an atypical, distributive reading of the basic maximizing insight that agents ought to create the most good they can. On that reading, agents do no wrong if, for each person who does or ever will exist at a world, those agents create the most well-being they can for that person at that world. Now, this principle provides only a sufficient condition for avoiding wrongdoing since, clearly, agents who fail to maximize well-being for some particular person have not necessarily done anything wrong.[2] Yet when a choice does manage to

[1] Derek Parfit, *Reasons and Persons* (Oxford, 1987), p. 363.

[2] The bare fact that agents act in a way that diminishes well-being for, or harms, a person does not imply that that act is wrong. After all, in conflict situations, agents may be forced to decrease well-being for one person in order to increase well-being for another. But we do not think that agents have wronged the one person, or have done

© 2007 Cambridge University Press
Utilitas Vol. 19, No. 3, September 2007
doi:10.1017/S0953820807002609 Printed in the United Kingdom

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

satisfy such a stringent sufficient condition – when it is a maximizing-for-each-person choice – we may well feel that we intuit that that choice cannot be wrong. Of course, the more typical, aggregative reading of the basic consequentialist insight rejects that intuition on the grounds that such a choice will still be wrong if it fails to maximize aggregate well-being – by, for example, failing to bring additional persons into existence. But that aggregative reading is itself in trouble.[3] While aggregative consequentialists have worked to overcome that trouble (by, for example, introducing other values as counters to aggregate well-being), even they do not claim to have finished the job.[4]

something wrong, if the tradeoff was made in some correct way. The plausibility of a person-based form of consequentialism will depend on whether plausible provisions governing conflicts can be formulated. If we find the basic maximizing insight attractive, then we would want to start out, at least, defining those provisions in a way that remains true to the spirit of maximization. Leximin, according to which agents must first create additional well-being for the least well-off, then for the next least well-off and so on, seems to meet this condition but remains highly controversial. For further discussion, see Roberts, 'A New Way of Doing the Best That We Can: Person-Based Consequentialism and the Equality Problem', *Ethics* 112 (2002), pp. 338–49. As Broome has shown, some ways of translating the person-based intuition into a consequentialist framework are problematic (John Broome, *Counting the Cost of Global Warming* (Cambridge, 1992), pp. 125–30). However, consistent formulations of the intuition within the context of maximizing consequentialism are also available. See Roberts, *Child Versus Childmaker: Future Persons and Present Duties in Ethics and the Law* (Lanham, Md.: 1998), pp. 45–85; and 'Is the Person-Affecting Intuition Paradoxical?', *Theory and Decision* 55 (2003), pp. 5–15.

[3] Traditional, aggregative consequentialism provides that it is not enough to maximize well-being for each person if the alternative exists of leaving those people as they are, or leaving them out of existence altogether, and creating additional people whose individual well-being levels make the greater contribution to total aggregate well-being. But it is, pre-analytically, odd to think agents are obligated to maximize the well-being of individuals en masse, including by way of constructing whatever mass of individuals it is that maximizes total aggregate well-being. Post-analytically, matters get still worse, since that same aggregative reading opens the door to a number of particularly penetrating criticisms. Thus, aggregative consequentialism arguably gives rise to (1) especially virulent forms of Parfit's repugnant conclusion problem; (2) Vallentyne and Kagan's infinite population problem; (3) implausible duties to procreate when we as individuals would rather not; and (4) equality problems that cannot reasonably be ignored. See Parfit, *Reasons and Persons*, pp. 381–90; Peter Vallentyne and Shelly Kagan, 'Infinite Value and Finitely Additive Value Theory', *Journal of Philosophy* 44 (1997), pp. 5–26; and Roberts, 'A New Way of Doing the Best That We Can', pp. 322–3.

[4] Some forms of consequentialism recognize a plurality of values in an attempt to overcome the deficiencies of traditional, aggregative consequentialism. Such theories take into account goods in addition to the good of aggregate well-being, including, for example, equality. See, e.g., Larry Temkin, *Inequality* (New York, 1993), pp. 222–7; Jeff McMahan, 'Wrongful Life: Paradoxes in the Morality of Causing People to Exist', in *Rational Commitment and Social Justice: Essays for Gregory Kavka*, ed. Jules L. Coleman and Christopher W. Morris (New York, 1998), pp. 208–47; John Broome, *Ethics out of Economics* (Cambridge, 1999), pp. 116–18; Broome, *Counting the Costs of Global Warming*, pp. 41–44; and Amartya Sen, *Inequality Reexamined* (Cambridge, Mass., 1992), pp. 7–8, 129–52. A concern is that any theory that holds that additional well-being in itself makes a significant contribution, however small, to the overall good will be vulnerable to extreme versions of the same old problems. Temkin, *Inequality*, pp. 224–5, accordingly,

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

On the non-consequentialist side, the person-based intuition flourishes. Thus, any approach that focuses exclusively on our reasons for acting as we do, or on those principles whose universal acceptance we all could rationally will, or on our virtue or character, or on individual rights, in ways that do not somehow make room for the consideration of what our choices do to others as individuals will encounter immediate resistance.

As these points suggest, the person-based intuition marks badly needed common ground between at least some forms of consequentialism and various non-consequentialist approaches. That an idea marks common ground hardly means that it is right. What is indisputable, though, is that, given the shortage, any otherwise plausible common ground we can find should be deemed worthy of our most serious consideration.

But in fact the person-based intuition has achieved no such lofty status. It has instead been scoured away by a handful of riveting future person cases collectively known as the 'non-identity problem'. For most consequentialists, the non-identity problem confirms that the person-based intuition must be discarded in whole or in part in favor of an impersonal (e.g. an aggregative) or a mixed competitor.[5] Non-consequentialists similarly have been driven to views that imply harm or wrong to persons who seem intuitively not to have been harmed or

has proposed that there may exist an upper limit on the contribution that additional units of well-being make toward the overall good. This capping proposal, though it runs counter to the idea that it is by virtue of some value that additional units of well-being independently have that they normally increase the overall good, seems potentially useful. The evaluation of it is, however, beyond the scope of this article. Some theorists, in addition, have proposed that the distinction between (1) whether a particular person's existing contributes to the overall betterness of an outcome and (2) whether, from the person's own point of view, the existence is worth having, can help to overcome some of those obstacles, in particular the repugnant conclusion. See, e.g., Broome, *Weighing Lives* (Oxford, 2004), pp. 210–14; and Fred Feldman, 'Justice, Desert, and the Repugnant Conclusion', *Utilitas* 7 (1995), pp. 189–206. But see too Feldman, *Pleasure and the Good Life* (Oxford, 2004), pp. 205–6.

[5] Consequentialist approaches (including those mentioned in n. 4) that work to avoid the challenges that face traditional, aggregative consequentialism take it for granted that the non-identity problem, as well, must be skirted. More generally, the impersonal, aggregative approach is considered to have the advantage of inherently avoiding the non-identity problem. Moreover, Temkin uses a version of Parfit's depletion example – what Temkin calls the 'live for today' problem – to undermine the 'leveling down objection' against the egalitarian position that he articulates. Thus, he argues that the leveling down objection is most persuasive for those who think that an option that is better for each person, while worse for none, is the morally better option in all morally significant respects. Since the non-identity problem undermines that person-based idea, the leveling down objection, Temkin argues, correspondingly loses force. See Temkin, 'Equality, Priority and the Levelling Down Objection', *The Ideal of Equality*, ed. Matthew Clayton and Andrew Williams (Basingstoke and New York, 2002).

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

wronged at all.[6] The person-based intuition has thus become a strictly minority view – if that.

In this article, I argue that the non-identity problem in one of its most challenging forms is itself mistaken. If there is a mistake, it is important that we recognize that fact. After all, it is not as though the non-identity problem is our only problem. We have other issues to deal with as well, in droves.[7] My concern is that, in continuing to work hard to avoid a seemingly gravely serious problem that isn't really a problem at all, while trying to construct a moral theory capable of addressing a range of gravely serious problems that do hold up to close scrutiny, we may be taking on an impossible task. Even if it is a task that we may be technically able to accomplish, there remains the danger that our

[6] The non-identity problem has also motivated some non-consequentialists to formulate a range of impersonal accounts of wrongdoing. Thus, an account that condemns acting 'wrongly toward' a particular person in some cases in which it is clear that agents have in every way acted for the good of that person, and with a good outcome for that person, is in a position to explain how it can be that an agent can do something wrong while technically wronging, or harming, no one. This suggestion was made by Kavka in response to the non-identity problem (Gregory Kavka, 'The Paradox of Future Individuals', *Philosophy & Public Affairs* 11 (1981), pp. 105–6). A related approach calls for the bringing into existence of a better-off person in place of a worse-off person in certain cases without suggesting that the worse-off person has been either harmed or wronged. See Allen Buchanan, Dan W. Brock, Norman Daniels and Daniel Wikler, *From Choice to Chance* (Cambridge, 2000), pp. 247–57.

Still other theorists have attempted to retain the semblance of the person-based approach while defining 'harm to persons', or alternately 'wrong to persons', in a non-comparative way, i.e. one that disconnects those notions with how well-treated or well-off a given person is, given the alternatives. One such proposal suggests that we deem a given person to have been harmed when the only means by which agents can increase well-being is to bring some better-off but distinct person into existence in place of that one. On that account, harms are avoidable 'by substitution' of one individual for another. See Philip G. Peters, Jr., *How Safe is Safe Enough? Obligations to the Children of Reproductive Technology* (Oxford, 2004), pp. 20–39. Peters' approach allows us to talk in person-based terms, but the theory that underlies that talk seems impersonal in nature. Others propose to simply define 'harm' to include pain, hardship or deficiency (beyond a certain threshold) even in cases in which well-being has been maximized and agents have no option for improving the individual's situation. See Elizabeth Harman, 'Can We Harm and Benefit in Creating?', *Philosophical Perspectives* 18 (2004), pp. 89–113, esp. p. 93; and Bonnie Steinbock, 'The Logical Case for Wrongful Life', *Hastings Center Report* 16 (1986), pp. 15–20. See also Cynthia Cohen, 'Give Me Children or I Shall Die! New Reproductive Technologies and Harm to Children', *Hastings Center Report* 26 (1996), pp. 19–27. A somewhat different tack proposes that a person can be wronged even in a case in which he or she has not been harmed. See Rahul Kumar, 'Who Can Be Wronged?', *Philosophy & Public Affairs* 31 (2003), p. 112; James Woodward, 'The Non-Identity Problem', *Ethics* 96 (1986), pp. 804–31; and Woodward, 'Reply to Parfit', *Ethics* 97 (1987), pp. 805 (defining the extent of a 'wrong' as the difference between a person's actual state and an 'unobtainable baseline' in which that person's rights are 'not violated').

[7] Aggregative consequentialists, for example, cannot rest content avoiding the non-identity problem. They must address other serious problems as well, including the repugnant conclusion. See n. 3.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

work product will itself be distorted and even mistaken in important ways.[8]

Close inspection reveals that the non-identity problem is really a collection of problems that have different logical features. Non-identity problems thus can be typed. Unsurprisingly, while all types of the non-identity problem purport to show that what is 'bad', or wrong, is in some cases 'bad for', or harms, no one, distinct types have distinct levels of potency. For example, one type of non-identity problem – the counterfactual problem; we can call it the '*won't*-do-better' problem – can be quickly unraveled once we reject the highly problematic counterfactual account of harm on which it relies. Suppose, for example, I shoot you in the arm and that, had I not shot you in the arm, I (being angry) *would* have shot you in the heart. This is not a case in which the act under scrutiny is not bad for and does not harm you. We will come back to the won't-do-better problem, briefly, in what follows, but for the moment it is enough to say that an account of harm that is unreliable in an ordinary existing person context cannot be made reliable when transferred to the trickier future person context.[9]

A second type of non-identity problem is a more serious threat to the person-based intuition. We can call it the '*can't*-do-better' problem

[8] If we do end up with theories that are too complex, vague, nuanced and indefinite to be assessed or applied, or are so narrow that our acceptance of them must be tentative pending an understanding of how they fit into a broader theory, there will be practical implications. For example, we would surely need to suspend hope that moral theory might have some advice to offer courts as they struggle to decide hard 'future person' cases in the law. And there is a deeper problem: a moral theory stripped of the person-based intuition will not even be cognizable within the law. The common law of torts and constitutional privacy law are inherently person-based; without a radical restructuring of the law, wrongs will not be recognized in the absence of a genuinely (not stipulatively) harmed person. Thus, constitutional principles protect privacy interests when the activity at issue harms no one and the state is therefore unable to justify regulation or prohibition; and tort law principles reject complaints in negligence, for failure to state a claim, in the absence of harm. Of course, we might completely revolutionize those person-based branches of the law. But we should be hesitant to advise that course, in view of the fact that impersonal alternatives, at this juncture, come with their own challenges.

On the other hand, if an important type of non-identity problem is indeed incorrect, then it is important to recognize that fact earlier rather than later. If we no longer have to tiptoe around that problem, then consequentialists, including pluralists, can revisit the aggregative, impersonal aspects of their views and address the repugnant conclusion and other challenges in new ways. The intuitive 'levelling down' objection against the ideal of equality will no longer be undermined by Temkin's 'live for today' example. See n. 5. For the non-consequentialist, the connection between 'acting wrongly toward' and wronging persons can be re-established. Harm-to-a-person can be left harm-to-a-person instead of giving way to the failure-to-substitute-a-better-off-person. And, finally, we will have no need to give up useful distinctions between imposing pain on and harming a given person, and between harming and wronging a given person. See n. 6.

[9] See sect. III. The shoot-you-in-the-arm-or-the-heart example derives from Feinberg's 'two schoolyard bullies' example. See Joel Feinberg, 'Wrongful Life and the Counterfactual Element in Harming', *Social Philosophy & Policy* 4 (1988), pp. 150–2. See also *Child versus Childmaker*, pp. 160–4.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

since it arises when a given person's maximal level of well-being, though unambiguously positive, is capped due to factors – for example, genetic factors – beyond any agent's control. To preserve the person-based intuition, we must come to the view that in some such cases – where, for example, the choice maximizes well-being for each existing and future person – it is morally permissible to produce an impaired child when one might have produced, at no real cost to anyone, a distinct, better-off child instead. But in the end I think we may be willing to take that step – *provided that* we feel sure that in doing so we will not be forced to accept the same result for a broader array of far more disturbing non-identity problems, including Parfit's depletion example. The challenge presented by the can't-do-better problem is thus not an obviously insurmountable one.[10]

And, finally, there is a third type of non-identity problem: the 'can't-*expect*-better' problem. It is that third problem that is at work in Parfit's depletion example, Kavka's slave-child case and the transgenerational compensation puzzles on which Sher focuses, as well as in many other cases, and it is that problem I focus on here.[11] My purpose is to show that in the relevant cases the clearly bad act after all is bad for someone. In making this argument, I will not quibble about whether the acts described as 'bad' in the problem cases are bad. And I will not rely on a merely stipulative notion of harm that is at odds with our intuitive, comparative concept of harm.[12] Rather, my target will be the line of reasoning that purports to take us from that intuitive, comparative concept of harm to the 'no harm done' result that is critical to the success of the can't-expect-better problem.

That line of reasoning seems ordinary on its face. Asked to compare a not-so-wonderful alternative for a given person against a vanishing probability of there being any better alternative, we of course favor

[10] See sect. VI.

[11] See Parfit, *Reasons and Persons*, pp. 361–4; Kavka, 'The Paradox of Future Individuals', p. 100; and George Sher, 'Transgenerational Compensation', *Philosophy & Public Affairs* 33 (2005), pp. 185–200. See also Kazuo Ishiguro's story of clones brought into existence only to serve as organ donors: *Never Let Me Go* (New York, 2005). The three types of non-identity problem I review in this article do not exhaust the kinds of non-identity-like arguments that have been put forward to show 'no harm done' to persons whose existence is in some way within the agent's control. Some of those arguments, however, raise no genuine question of harm at all and should simply be set aside. See n. 54.

[12] Parfit and Feinberg both retain an intuitive, comparative, person-based criterion of harm. Thus, Parfit writes that acts that are 'clearly better for' a given person do not harm that person in a 'morally relevant' sense (*Reasons and Persons*, pp. 373–4). Feinberg likewise argues that harm to a person depends on the comparison between what things are in fact like for that person and a certain 'baseline' ('Wrongful Life and the Counterfactual Element in Harming', p. 147). Other theorists reject a comparative approach. See n. 6.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

the former. Inferences made on the basis of such comparisons are often cogent. A bird in the hand is worth two in the bush. But in other cases slight distortions can generate lines of reasoning that are not cogent. I argue, in particular, that the can't-expect-better problem makes not just one extremely subtle mistake but rather alternates between two. We sort out one fallacy, in other words, only to fall unsuspecting and against all odds (as it were) into a second.

I have elsewhere discussed various types of non-identity problems.[13] Those discussions have been, however, schematic in certain ways. By here focusing more clearly and in more detail on one account of harm in particular – an account designed to reflect the role that a well-defined concept of expected well-being plays in establishing harm – and the rules that must govern that account, I hope to provide a more compelling argument that the can't-expect-better problem can be set aside.

* * *

Section II below describes the can't-expect-better problem. Section III outlines the deficiencies of a counterfactual account of when a choice is bad for, or harms, a particular person. I argue in that section that harm is correctly assessed by reference not to some narrowly restricted set of alternatives – those alternatives, for example, agents happen to be willing to perform – but rather to a considerably broader set of alternatives. I urge, in other words, that we adopt a *maximizing* approach to harm. An *objective* maximizing account quickly reveals its limits. In contrast, an *expectational* maximizing account seems more plausible. But applying that account can be tricky. A simple lottery example, described in section IV, is useful for identifying the fault lines.

The lottery example is interesting in its own right. For our purposes here, however, the example is important because the line of reasoning that generates the 'no harm done' result in the lottery example parallels the line of reasoning that we see in the can't-expect-better problem. That means that, if we can see clearly why the inference fails in the lottery example, we should be able to see why the inference fails in the can't-expect-better problem as well.

The depletion problem itself is the focus of section V, where I argue that we can after all discern how the agents of depletion have in fact

[13] See *Child Versus Childmaker*, pp. 87–133; and 'Is the Person-Affecting Intuition Paradoxical?', pp. 21–34 (focusing on Kavka's 'slave-child' example). I have also discussed the non-identity problem as it arises in connection with some variations on the repugnant conclusion. See 'Person-Based Consequentialism and the Procreation Obligation', *The Repugnant Conclusion: Essays in Population Ethics*, ed. J. Ryberg and T. Tännsjö (Dordrecht, 2005), pp. 99–128.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

acted to harm people still a long way from existence at the time of the performance of the act under scrutiny. There are two tricks for discerning harm in such cases. We must accept that *all* alternatives open to the agent at a given time, not just some narrowly restricted set, are pertinent to the issue of harm. And we must take care to avoid at least two beguiling errors in the application of the otherwise plausible expectational account of harm.

Conclusions are drawn in section VI. I concede there that any progress we make toward solving the can't-*expect*-better problem leaves the can't-*do*-better problem untouched. For the person-based theorist, then, the can't-do-better problem emerges as the greatest challenge.

II. THE CAN'T-EXPECT-BETTER PROBLEM

The can't-expect-better problem appeals to Kavka's observation that the coming into existence of any one particular person ordinarily is highly 'precarious'.[14] In other words: for any particular person, the odds, at least until conception, are ordinarily stacked overwhelmingly against existence.

Kavka's own slave-child case is one of the most compelling in this genre.[15] In that case, intuitively and by hypothesis, a couple who enter into a contract to sell their future child into slavery – a child who is then born and suffers as a result of his or her status as slave, yet has a life that is unambiguously worth living – have done something wrong. This is so, despite the fact that any change in the actual sequence of acts and events – including the couple's not entering into the contract but still proceeding to conceive a child – would very likely have affected the timing and circumstances of conception, including the matter of which sperm fertilized which egg. Any such change thus very likely would have served only to take that original child off-track for existence altogether. Thus, perhaps the couple could have produced another child better off than the original. But that other child would also very likely have been 'non-identical' to the original child. Hence, producing a better-off child would not have made things better for the original child. If anything, it would have made things worse. From those points the result is drawn that the couple's entering the slave-child contract is not bad for and does not harm the child that it produces as a slave. But if the act is wrong but harms neither the child nor (also by hypothesis) anyone else, then the person-based intuition must itself be rejected as false.

[14] Kavka, 'The Paradox of Future Individuals', pp. 93–112.

[15] Kavka, 'The Paradox of Future Individuals', p. 93. See also Doran Smolkin's discussion of the 'Acme chemical company case' in 'Toward a Rights-Based Solution to the Non-Identity Problem', *Journal of Social Philosophy* 30 (1999), pp. 195–6 (citing James Woodward, 'The Non-Identity Problem', *Ethics* 97 (1986), p. 813).

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

It may seem that a maximizing approach has no particular ability to establish harm in the context of the can't-expect-better problem. Under such an approach, agents who have in fact done the best they can for a given person have avoided imposing any harm on that person. But it is exactly that maximizing feat that agents have accomplished, according to the can't-expect-better problem, when they bring a person into an existence that is, though flawed in some way, better than what agents *can* achieve, or what they can *expect* to achieve, for that person under each alternative act. For on close inspection each such act – including those that, on their face, seem to meet the most stringent standards – would have served only to create for that person no more than an extremely low chance of ever coming into existence at all. Even the substantially flawed existence – as the can't-expect-better problem so beguilingly observes – is better than, or at least not worse than, *that*.

Kavka's slave-child example is compelling. But because of its painfully true-to-life quality so is Parfit's depletion example.[16] We face pressing issues regarding the conservation of resources and the protection of the environment. If the person-based intuition implies that it matters not at all how we decide those issues in cases in which the horrible effects of depletion and despoliation will not be felt for many decades or centuries – at which point those who then exist and suffer will in some sense 'owe their very existence' to those choices – then we must reject that intuition.

III. WHEN A CHOICE IS BAD FOR, OR HARMS, A PERSON

A. A counterfactual account of harm

The counterfactual, or '*won't* do better', problem provides an excellent motivation for a maximizing account of harm. Suppose a couple, themselves genetically impaired in some serious way, want to have a child who is likewise impaired. Perhaps, for example, they are deaf.[17] Suppose that one of the woman's eggs has already been fertilized in vitro using her partner's sperm, and that a 'safe and effective' technology exists that would allow a doctor to introduce a gene for deafness into the genome of that developing embryo. The couple convinces the doctor to do just that. The baby is born, deaf, nine months later. Suppose that the baby would have been significantly better off had

[16] *Reasons and Persons*, pp. 361–6.

[17] If the assumptions of this example, including the assumption that the choice is wrong, do not seem plausible for the case of deafness, one can adjust the case, so that the impairment is not deafness but rather cystic fibrosis, Huntington's disease, etc. – any genetic or other disorder, including, for example, impairments associated with aggressive uses of fertility therapies leading to supernumerary pregnancy, where the life that results, assuming that the care that is owed is provided, is unambiguously worth having.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

he or she been born without that impairment. Suppose that – in part because the cost to the parents of having non-deaf offspring is either not grave or can be mitigated – what the couple has done is wrong. Suppose, finally, that, having a deaf child is so important to the couple that, had they not taken steps to have the preferred gene introduced into the embryo, they *would* have had that embryo destroyed, in which case their deaf baby would never have existed at all. The bottom line, then, for *this particular baby* is just this: the couple *won't* do better for *this baby* than they have, even in the case where they later decide to produce still another ('non-identical') baby who can hear. If anything, they would only do worse.

If on these facts we think that the couple's making the baby deaf does not harm the baby, then we confront a non-identity problem: a case where a wrong (by hypothesis) has been done even though no one has been harmed. However, as the 'shoot you in the arm or the heart' case shows, the counterfactual test is unreliable as a necessary condition of harm.[18] In fact, any close inspection shows us exactly how the couple has harmed their baby. Regardless of whether it is an alternative the couple themselves find appealing, the fact is that the alternative existed for the couple of both *not* 'knocking in' the gene for deafness and still bringing that particular baby into existence.[19] Now, to say that the alternative existed for them is not to say that the couple was *obligated* to have chosen that alternative. Nor is it to suggest that they *would* have taken that option, or would have been *inclined* to take that option, had they somehow been barred from pursuing the option they in fact pursued. Rather, it is the bare existence of that better-for-the-baby alternative that plausibly establishes harm in the case where the couple both brings the baby into existence and makes the baby deaf. Likewise

[18] This objection to the counterfactual test, which leaves open the use of the counterfactual test as a sufficient condition for harm, is described by Feinberg. See n. 9. Theorists sometimes use counterfactuals in describing how they reach a 'no harm done' result in the context of the non-identity problem. But often the language is probably shorthand for the sort of expectational account of harm described below. Kavka, however, expressly brings probabilities into the picture ('The Paradox of Future Individuals', p. 93 and esp. p. 100 n. 15). Other theorists mix counterfactual language with what seems at base to be the sort of objective account of harm described in sect. III.B below. See Smolkin, 'Toward A Rights-Based Solution to the Non-Identity Problem', pp. 195–6. There, Smolkin considers, as an example of a case in which the agent performs an act that is 'necessary' for another's existence, the Acme chemical corporation case. He then writes that, had the plant not continued in operation, A and B 'would never have met' and C 'would never have existed had the corporation not dumped the chemicals'. See also John A. Robertson, *Children of Choice: Freedom and the New Reproductive Technologies* (Princeton, 1994), pp. 121–2 and 168–9 (applying different types of the non-identity argument in the context of new reproductive technologies); and 'Procreative Liberty and Harm to Offspring in Assisted Reproduction', *American Journal of Law and Medicine* 30 (2004), pp. 13–17 (applying a more clearly maximizing approach to the issue of harm).

[19] Issues of personal identity do not arise in this context except for those theorists who accept a highly questionable genetic identity theory of personal identity.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

in the shoot-you-in-the-arm-or-the-heart case: it is the bare existence for me of the alternative of not shooting you at all that grounds the assessment that I harm you when I shoot you in the arm.

B. An objective maximizing account of harm

A maximizing account assesses harm based not on what the agent would have done but rather on what the agent could have done on behalf of a person. But the array of genuine alternatives open to an agent or group of agents at a given time is, while extensive, not unlimited. It excludes acts that secure outcomes that are themselves not *accessible* to the agent at that time – outcomes, or possible futures or worlds, the achievement of which is blocked by either natural law or the acts of other people that the agent has no ability to control or alter. We thus can say that:

> *Objective account*: A person p is harmed by an agent's performance at a time t and a world w1 of an act A *if and only if* p exists at or after t at w1 and there exists a distinct world w2 such that (1) w2 was accessible to the agent prior to t; (2) the agent performs an alternative act B at t at w2 instead of A; (3) p has more well-being at w2 than p has at w1; and (4) the agent's performing A at t at w1 and B at t at w2 are parts of the causal sequences at w1 and w2 that end (respectively) in p's having the well-being that p has at w1 and in p's having the well-being that p has at w2.[20]

[20] The objective account provides an account of when an agent *as an individual* harms a person. Agents can also impose harm not as individuals but rather *as participants in a group*. The concept of group-harm becomes particularly important in addressing the problem of 'causal overdetermination'. For our purposes here, however, we will assume that our non-identity cases (e.g. the depletion example) are not complicated by that further problem. We will thus assume that an outcome that is accessible to a group of agents (e.g. the present generation) is also accessible to each agent as individual. For further discussion, see Roberts, 'Supernumerary Pregnancy, Collective Harm and Two Forms of the Non-identity Problem', *Journal of Law, Medicine & Ethics* 34.4 (2006), pp. 776–92.

According to the objective account, people who are merely possible at a world w1 – people, that is, who do not and never will exist at w1 – cannot be harmed at w1. Another option would be to accept that not bringing a person into existence may harm that person but then to stipulate, in our account of when a person is *wronged*, that any person who does not and will not exist at w is not wronged at w.

The objective account does not require that p exists at the time the act itself is performed; if it did, we would immediately lose the potential to address any form of the non-identity problem. Nor does it require that p exist, at some time or another, at the distinct world w2 at which the agent creates additional well-being, since we want to be able to say that an agent in some cases harms a person by failing to *avoid* bringing that person into existence – as when, for example, the life is, as a function of a genetic or chromosomal defect, one that is less than worth living.

I leave the causal relation referred to in clause (4) unspecified. But it is intended not to import a counterfactual test of causation. Its purpose is rather to insure that an act A that in fact does not harm a given person p is not deemed to harm p by virtue of the fact that some second act is performed that, *despite* A, really does make things worse for

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

This account of harm, which asks simply whether agents could have created more well-being for a particular person than they have, is a maximizing account of harm. And it is an objective account of harm in the sense that the test is whether there as a matter of fact existed for the agent a way of bringing about a better-for-p outcome, regardless of whether the agent is in a position to identify precisely what he or she needs to do in order to achieve that outcome for p.[21]

The objective account gives us a way of addressing the can't-expect-better problem. After all, there is nothing in natural law or the acts of other agents that bars the couple who produce the child as a slave from producing the very same child as a non-slave – or that bars agents from implementing conservation over depletion in a way that improves the plights of at least some members of the future generation who exist and suffer under depletion.

Now, it may be objected that the objective account supports an unnaturally broad use of the term 'harm'. I am not sure, however, that that broad usage is not entirely natural so long as we understand

p. We need not dwell on these technicalities, however, since the cases of interest here all easily meet the condition. Thus, the couple's having entered into the slave-child contract is, at the 'actual' world, part of a causal sequence that ends in the child's being born a slave; likewise, the couple's taking a sip of water instead of entering into the slave-child contract is, at a distinct accessible world, part of the causal sequence that ends in the child's being born a non-slave; etc.

Whether a given person will exist or not and be harmed or not as a consequence of a given act's being performed will often depend on how the future otherwise unfolds. To insure cogency, both the objective and the expectational account that follows relativize inferences of harm to worlds, where, for any given existing or future person, it is a determinate matter of fact whether that person exists, at least eventually, or not, and if so how things turn out for that person. Nothing said here, however, implies that agents can make the inference, at or just prior to t, that an act performed at t harms any *particular* person who does not exist at t, since, at t, agents are in no position to know, for any particular person, that he or she will exist at some future time and then suffer. Indeed, it may not even be a determinate matter of fact at t for any particular person who may be affected one way or another by the act under scrutiny that that person will exist. It all will depend on how the future unfolds. On the other hand, agents do not need to make any such inference for any particular person in order to consider whether, or know that, what they are about to do will harm some person or another. They, after all, can figure out perfectly well that, if they do A, then whether the future unfolds in one particular way rather than another, so that one collection of people rather than another exists, *someone* will be harmed.

Woodward explicitly appeals to a maximizing approach in many of the variations on the non-identity problem that he considers. But he then unaccountably declares the risky policy to be a 'necessary condition' for the existence of those people who exist and suffer hundreds of years later and describes the situation in which any of them are made better off as an 'unobtainable baseline'. See 'The Non-Identity Problem', pp. 812 and 817–18. At points, Parfit seems to concur. See 'Comments', *Ethics* 96 (1986), p. 855.

[21] In describing the account as 'objective', I mean to indicate that it is not probabilistic and not (just) forward-looking. This is not, however, the best term, since as I understand it the expectational account as well can be developed as an account that is itself objective and not merely subjective in nature.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

harm to be necessary but not sufficient for a person's being wronged or for wrongdoing generally.[22] Thus, perhaps I do *harm* Bill Gates when I fail to endorse my paycheck over to him each month. But we so quickly and automatically conclude that I nonetheless do not *wrong* him, or necessarily do any wrong at all, when I use my money for other purposes that we never find ourselves working hard to block an obviously problematic inference.[23]

However, the objective account raises more serious concerns as well. Imagine that the woman standing next to you on the subway suddenly collapses. Imagine that her collapse is due to heart failure but that, if you were to proceed in an exactly 'right' way, step by step, you could perform (impeded neither by natural law nor by the acts of other agents) then and there a successful open-heart procedure using just your pocketknife, saving the woman's life and her every function. This is so, despite the fact that your medical knowledge is almost non-existent and you have no clue even how to commence any such correct procedure.

The objective account implies that you harm the woman when you fail to perform a correct procedure. But that result may seem implausible. The reality is that, if you were to attempt to do what the objective account says you must do to avoid harming the woman – if you were to get out your pocketknife and go to work – it is highly probable that you would kill her. Plausibly, you can avoid harming the woman in this case not exclusively by performing a correct procedure but also by shouting for a doctor and beginning CPR – even in the case where doing the latter rather than the former in fact creates less well-being for the woman. Plausibly, we need an account of harm that confirms that intuitive fact. Such an account of harm would make harm the kind of thing we often know how – step by step – to avoid imposing, should we choose to avoid imposing it, and thus would accord to harm the action-guiding potential that we intuitively think facts about harm surely often can provide.

[22] We are, after all, accustomed to relativizing, implicitly or explicitly, the sorts of comparisons that would support a finding of harm to a limited number of alternative scenarios that we think are important for one reason or another and then using the term 'harm' to rank those scenarios in a way that seems natural. For example, it seems natural to say that a particular beneficiary of Save the Children has been harmed compared to how well off he or she would have been had many people given more than they have to that organization. But if that use of 'harm' is legitimate, then it seems that the sense of 'harm' generated by the objective account may be in line with ordinary usage after all. We can think of that latter sense of 'harm' as the existentially quantified version of the former, comparative claim: the child is harmed because he or she is harmed compared to how well off he or she would have fared had people been more generous.

[23] I owe this example to David Wasserman.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

C. An expectational maximizing account of harm

Under the objective account, an implication of harm arises in virtue of the fact that there exists an outcome that is accessible to the agent and better for the individual and despite the fact that that outcome is one the agent in a million years could never achieve. A better account of harm might seem to be one that would assess harm not just by reference to the constraints imposed by natural law and the acts of other agents but by reference to the various *epistemic constraints* under which agents must operate as well. Such an account would recognize the action-guiding function that we intuitively think that facts about harm can potentially have by generally making the matter of whether a particular alternative act contemplated by the agent will harm someone or not something that the agent can detect *prior* to his or her performance of that act. On the basis of that information, the agent can then choose to go forward with the contemplated act in place of some other or not. Whether harm is done or not, then, would remain something that is generally within the agent's own control.

An account that understands harm in this way will require for harm that the information necessary to decide whether a particular act will impose harm or not is information that the agent is in a position to grasp *at that critical time just prior to performance*. Now, the sort of information that agents are typically in a position to grasp at a time at which they have not yet implemented any particular choice comes in the form of connections between (a) the acts that exist as alternatives for the agents at that time and (b) the results, or outcomes, that we may be led to at least in part as a result of those acts being performed. Such connections are themselves *probabilistic* in nature. Thus, in the collapsed subway rider example, you are well-positioned, prior to performance, to grasp that your making an incision to commence an open-heart procedure will probably, in view of the untutored nature of that incision, kill the woman. If you luck out at step one and she survives, there is always step two. What you are in no position to know is how each of the many distinct alternatives for starting an open-heart procedure that exist for you prior to performance is connected with a particular outcome for the woman – which alternatives, in other words, will commence a correct open-heart procedure and which will commence a murderous sequence of acts and events that will end with the woman's death. Since the information must be had (or at least must be graspable) prior to performance, it is the *probabilistic connections* that this second account of harm must exploit – and the 'actual' connections, the connections you are able to have no real knowledge of prior to performance, that it must set aside.

It is, of course, exactly such probabilistic connections that aggregative consequentialists appeal to when they shift from *actual* to *expected*

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

value in order to take into account the reality of epistemic constraints.[24] And, indeed, the usual rules for calculating expected value can easily be adapted for person-based purposes – that is, for the purpose of calculating the expected well-being *for a particular person* that is generated, and ultimately the harm that is imposed *on that person*, by the performance of one act rather than some alternative. Thus, where the act under scrutiny is to be performed at time t, to determine the expected well-being that that act creates for a person p, we can add up, possible outcome by possible outcome, the actual well-being that p has at that outcome multiplied by the probability, calculated at a time t^* just prior to t, that that outcome will obtain given the performance of that act.[25]

Of course, as time marches on and the epistemic situation evolves, expected well-being may itself vary. But for purposes of the expectational account it is critical that the probabilities themselves are calculated on the basis of information that the agent is in a position to grasp *at that time just prior to performance* regarding both the coming act itself and how the future that that act will fit into will otherwise unfold. For it is *at exactly that critical time* that the agent both (a) is in the best epistemic position that he or she can be in regarding the likely effects of the act under scrutiny yet (b) still has the option of performing that act or not. At any *earlier* time, additional information may still reveal itself. But especially important for our purposes here is that the relevant probabilities cannot be correctly calculated at any time *later* than just prior to performance. A requirement that the probabilities be calculated at any *later* time would destroy the action-guiding potential that is the whole point of the expectational account. If the assessment of harm can be made only after the performance commences or after it is complete, then that assessment will fail to give the agent timely guidance on how to avoid imposing that harm.

Timing is, then, everything, under an expectational account of harm. Now, as of that critical time just prior to performance, the body of information available to the agent may well be highly restricted. This is so, in virtue of the fact that, prior to performance, many details

[24] Kagan describes a 'subjective account of rightness' that may help to respond to the objection that consequentialism is not useful. See Shelly Kagan, *Normative Ethics* (Boulder, 1998), pp. 64–9. A number of theorists outline how probabilities and outcomes can be put together to define 'expected utility'. See, e.g., Mark Timmons, *Moral Theory: An Introduction* (Lanham, Md., 2002), pp. 122–6; Graham Oddie and Peter Menzies, 'An Objectivist's Guide to Subjective Value', *Ethics* 102 (1992), pp. 512–33, esp. pp. 513–16; and Frank Jackson, 'Decision-theoretic Consequentialism and the Nearest and Dearest Objection', *Ethics* 101 (1991), pp. 462–7. Feldman critically discusses the 'epistemic objection' to an 'objective' theory of utility in 'Actual Utility, the Objection from Impracticality, and the Move to Expected Utility', *Philosophical Studies* 129 (2006), pp. 49–79.

[25] For these purposes, I assume that an outcome at which the person never exists assigns to that person a zero level of well-being.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

regarding just how the future will unfold – including details about how the contemplated act itself will unfold – remain obscure to the agent, who has no crystal ball. Indeed, prior to performance, it may not even be a determinate matter of fact that the future – or the contemplated act – will unfold in one particular way rather than another, even given that the agent makes one sort of choice rather than another. Going forward, these points will become critical. But they do not create any particular obstacle, under an expectational account, for discerning harm. Harm will simply be a matter of how much expected well-being the agent is able to create for a particular person, given the information about the future the agent is able to grasp at that critical moment just prior to performance, however highly restricted that body of information happens to be, by the performance of one act in place of some alternative act.

We can sum up the expectational account of harm as follows.

> *Expectational account*: p is harmed by an agent's performance of an act A at t at w1 *if and only if* p exists at or after t at w1 and there exists a distinct world w2 such that (1) w2 was accessible to the agent at a time t^* just prior to t; (2) the agent performs an alternative act B at t at w2 instead of A; and (3) the expected well-being of the agent's performance of B, calculated at t^* at w1, for p is greater than the expected well-being of the agent's performance of A, calculated at t^* at w1, for p.[26]

This principle easily generates the desirable 'no harm done' result in the collapsed subway rider case. Your calling for help and starting CPR clearly creates more expected well-being for the woman than your putting your scant medical knowledge to work to begin an open-heart procedure. You thus avoid harming the woman when you take the former option rather than the latter, whether or not the woman's life and every function is in point of fact eventually saved by emergency rescue workers.

As we will very shortly see, we may want to narrow this account in a certain way. But I want first to note that this test sets very stringent standards for how agents must comport themselves to *avoid* imposing harm. For one thing, the expectational account is itself a maximizing account.[27] Under this account, your just sitting on the subway and watching the woman die would not itself be a way to avoid harming her. Second, while this account is consistent with a subjective, or an agent-relative, theory of expected well-being, it makes reference not to

[26] I am grateful to a referee for this journal for pointing out the need to include the requirement that 'p exists at or after t at w1' in the expectational account.

[27] Thus, modifying the objective account of harm in favor of an expectational account does not resurrect a counterfactual approach to harm.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

what the agent (who may willfully grasp little) in fact grasps but rather to what he or she is in a position to grasp at the critical time.[28] The agent's saying, or believing, 'I hadn't a clue how to go about performing CPR or whether it might help' thus does not on its own imply 'no harm done'.

Just as in the case of the objective account, the expectational account will no doubt seem too broad unless we understand that harming a person is not, on its own, sufficient for wronging that person or for wrongdoing generally. But even with that understanding, we may think the account is too broad. For we may think that harm is not a matter merely of a risk being imposed – which is what happens when agents create less expected well-being for a person when they have the option of creating more – but also of that risk eventuating. We may think, in other words, that harm must involve some sort of objective element, so that an act that does not lead to anyone's actually being made worse off is a 'miss' that, morally speaking, 'doesn't count'. If we think that, then we are free to revise the expectational account to include such a further necessary condition on harm. We can even adapt for that purpose the objective account of harm that we earlier rejected, and take the view that that which prospectively looks like it will harm someone but that as things turn out maximizes actual – not just expected – well-being for that person is not really a harm at all. If you begin the open-heart procedure with your pocketknife and don't instantly kill the woman, and a real cardiac surgeon, horrified by the prospect of your continuing, is driven to step in and complete the procedure, saving the woman's life and her every function, we can then say, plausibly, that your 'miss' doesn't 'count' and does not, after all, impose harm.[29]

Even as amended, the expectational account may still seem to be somewhat broad simply by virtue of the fact that it is a kind of

[28] Several issues of clarification arise at this juncture. Should the agent be held to a 'reasonable person' standard, or an 'ideal person' standard, or some still more objective standard? See Oddie and Menzies, 'An Objectivist's Guide to Subjective Value', p. 517 (arguing that a moral theory cannot 'provide a practical guide to action' 'without according a privileged position to the notions of objective chance and, by virtue of that, objective value'). Should we, moreover, understand the 'information' that the agent is in a position to grasp at a given time to exclude inaccurate information? For purposes of this present article, I put those issues aside.

[29] An interesting issue is how much harm is created – the very small amount that is represented by the difference between the expected well-being in fact created for p and the maximal expected well-being that could have been created for p, or the more significant amount represented by the difference between the well-being created for p and maximal actual well-being that could have been created for p. I will not try to resolve this issue here, but rather just note that we are free to appeal to some appropriately limited objective element as a measure of the amount of harm actually done, and to understand expected well-being to represent the causal element in harming that Feinberg noted. See n. 9 above.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

maximizing account. However, for purposes of addressing the can't-expect-better problem, that the account is broad in that particular way is an advantage. If the implication of 'no harm done' could be avoided simply by a shift from a narrow and more conventional account of harm to a maximizing account – a shift many consequentialists would be naturally inclined to make and non-consequentialists are perfectly free to make as well – then the can't-expect-better problem would lose its sway. The reason it is such a powerful problem, in other words, is just that the underlying account of harm it exploits is itself plausibly broad yet still seems to lead to the 'no harm done' result.[30]

IV. THE LOTTERY EXAMPLE

A. An existing-person variation on the can't-expect-better problem

I want to claim that Parfit's depletion example is exactly analogous at relevant points to the lottery example that I describe in this section. They are both, in other words, instances of the can't-expect-better problem. One difference is that the classic non-identity problem challenges the claim of harm in the case of *future* persons while the lottery example will turn out to be a strictly *existing-person* case. I will argue, however, that the existing-person versus future-person

[30] In part to give credit where credit is due and in part to suggest that the concept of harm isolated by the expectational account is not *sui generis*, I note that similar probability issues arise in negligence law (specifically, in medical malpractice) and that some courts have accepted claims of harm when those issues arise even while forced to operate within the person-based framework on which the law is premised. See *Herskovits v. Group Health Cooperative of Puget Sound*, 664 P.2d 474, 475 and 487 (Wash. 1983) (medical malpractice case in which the plaintiff claimed that the defendant had failed to promptly diagnose lung cancer, reducing the decedent's chances of survival from 39 to 25 percent). The majority opinion in *Herskovits* treated the question as one of causation rather than harm, finding that the defendant's conduct was a substantial factor in the death. However, a concurring opinion treated the issue as one of harm. While the concurrence did not suggest that the harm for which the defendant should be held liable was the whole of the decedent's death, it is nonetheless significant that the court recognized the presence of an 'actionable injury'. See also *Kallenberg v. Beth Israel Hospital*, in which the court allowed the jury to consider the issue of the defendant's liability for the death of the decedent in a medical malpractice case in which the decedent, the victim of a series of strokes, would have had, had she been properly treated, a '20, say 30, maybe 40% chance of survival' (45 A.D.2d 177, 179 (1975) (holding that the jury would be permitted to consider that the decedent '*might* have improved sufficiently, even after August 22, to undergo surgery and make a recovery'; emphasis added). Other courts have been similarly reluctant to accept the 'doomed either way' defense against a claim of harm and have favored a 'loss of chance' approach. See, e.g., *Jorgenson v. Vener*, 2000 S.D. 87, P 17, 616 N.W.2d 366, 371 (2000). The 'loss of chance' doctrine has also been suggested as a possible correction to the U.S. Supreme Court's treatment of statistical evidence in *McCleskey v. Kemp*, 481 U.S. 279 (1987). See Henry Louis Gates, in *Statistical Stigmata in Deconstruction and the Possibility of Justice*, ed. Durcilla Cornell (New York, 1992), p. 338 (citing notion of probabilistic harm described by K. A. Appiah, unpublished manuscript, 1989).

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

difference is in the end not material. Thus, if we think we can unravel the line of reasoning that seems to lead to a 'no harm done' result in the lottery example, we should be able to do so in the context of the depletion example as well.

Suppose, then, that William has bought a ticket for a lottery in which balls are drawn from a bubbling ball-bouncer to determine a winning number. On each ball is written a single digit 0 through 9, with digits evenly distributed. The winning number consists of a thousand digits, and the jackpot a billion dollars. The woman standing before the cameras who is to draw the balls – the agent – first pauses for just a few moments to steal half the jackpot by ordering, via her BlackBerry, the transfer of that amount to a numbered Cayman Islands account. She then manually begins to draw the thousand-digit number, first plucking an '8' from the bubbling bin, then some other number, then some other number, and so on. Lo and behold, a few hours later, William turns out to have the winning number, and he collects what remains of the jackpot.

We need to make two further assumptions. First, the stealing itself is wrong.[31] And, second, no one other than William is harmed, or wronged, by that stealing.[32] If, given those assumptions, the expectational account implies that the agent's stealing does not harm William, then we have in the lottery example a perfectly good, present-person, non-identity argument against the person-based intuition.

B. Calculating expected well-being under a non-stealing alternative

Did the agent's pausing for a few moments to steal half the jackpot via her BlackBerry *harm* William? Pre-analytically, it may seem that

[31] Thus we assume that, if one is going to run lotteries, one ought not run them in the way the lottery is run in the current example. (Given the size of the jackpot, it might be noted that a lottery that gives the holder of the ticket a one in 10^{1000} chance of winning a billion dollars is itself pretty unfair unless the tickets are available at a *very* small cost. But we set that point aside for purposes here.)

[32] Obviously, others may have been harmed, or wronged, in addition to William. Thus, losing players hold tickets that are, in effect, worth transiently less than they should be during that period between the agent's stealing the money and the selection of a number other than their own. But any such loss is irrelevant to the overall argument of the present article, given the analogy between the purchasers of losing lottery tickets on the one hand and those persons who could have, but never in fact do, come into existence on the other. We can take for granted here that those who never exist at all cannot be harmed, in any morally significant sense, or wronged. One might think that one person who surely has been harmed by the agent's act is not William but rather the person who 'would have won' had the agent opted not to steal. The difficulty with the application of that (counterfactual) test is that no one person 'would have won' had the agent performed some non-stealing alternative and then proceeded to draw some winning number or another. It all depends on just how the agent refrains from stealing and moves step by step through the draw.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

it did. And that is what the objective account of harm implies.[33] After all, nothing prevents the agent – no natural law; no other agent – from creating still more well-being for William by not stealing and then drawing the identical number – William's number – that she in fact draws. But we have now said that the objective approach is more plausibly adapted as a mere necessary condition for harm. While it is nice to know that that necessary condition is satisfied in the lottery example, the harder question is whether an act existed as an alternative for the agent at the critical time that would have created more *expected* well-being for William than the agent's stealing does.

A sampling of plausible candidates for such an act would seem to include the agent's using her BlackBerry to make a contribution from her own personal account to Save the Children, the agent's simply standing there at the dais taking deep breaths and the agent's feigning to use her BlackBerry to steal. Since these seemingly better-for-William alternatives clearly exist for the agent at the critical time, it may appear that the expectational account safely implies that the agent's stealing harms William. But one who thinks this has 'overlooked', as Parfit puts it, the non-identity problem.[34] One has failed to note that Kavka's point about the precariousness of existence extends to the precariousness of future events generally. In the lottery example, the stealing and William's subsequent winning have, by hypothesis, already taken place. For William to win under a non-stealing alternative, the agent must draw just those thousand digits that she in fact has drawn without any variation in order. But surely the agent's substituting in a non-stealing alternative for her actual stealing will make it highly likely that exactly such a variation will occur. After all, even the smallest change in the agent's timing or technique in plucking the balls from the bubbling bin is likely to result in some variation in which digit is drawn at which point along the way. She will pluck a '7' to begin with instead of the critical '8', or if she lucks out with the '8' the next digit will be off. And the shift from stealing to non-stealing hardly counts as a small variation. It seems more an event of seismic proportions. Of course, if the agent does not steal, someone other than William – someone 'non-identical' to William – may well win a still bigger jackpot than William has in fact won. But that sequence of events will do William no good at all – and, indeed, will leave him worse off than he in fact is given the agent's stealing.

It is clear, then, that under a non-stealing alternative it is extremely unlikely – provisionally, it is something like 1 in 10^{1000} – that William will win the lottery. Of course, had the agent performed a non-stealing

[33] This assumes, of course, an implausibly direct connection between money and well-being.

[34] *Reasons and Persons*, p. 367 (writing in the context of the depletion example).

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

alternative and had William, against all odds, nonetheless won the lottery, he would have been substantially better off. That fact obviously must be reflected in the calculation of expected well-being. However, when the chances of winning are as low as William's are here, and the chances that William will be left with the minimal well-being level that he starts with as high as they are here, even substantial actual well-being translates into a level of expected well-being that is only barely there.

It might be objected that this initial estimate of William's chances of winning under a non-stealing alternative cannot be exactly accurate. It fails to recognize that the agent has certain strategies available, consistent with not stealing, that would boost somewhat William's chances of winning above that base level. Thus, the agent may grasp, prior to performance, certain features in her yet-to-be-performed stealing – features that will help to decide who will win; features that will, indeed, boost to some extent the chances that the winner will be William. The agent can then engineer, or *replicate*, those features in a non-stealing alternative. How can she manage to achieve that result? How can she engineer features of her yet-to-be-performed *stealing* in a *non-stealing* alternative? The trick is that the very sorts of features that help to decide who will win in the end have nothing to do with the actual act's being an act of stealing.[35] Being an act of *stealing* is obviously not a feature that can be replicated in a non-stealing alternative. But nor is it the sort of feature that on its own influences the causal order in such a way as to help decide who wins. Rather, the sorts of features that perform that latter function are the various *spatial-temporal-mechanical* features that the agent's actual stealing will turn out to have, whether by design or otherwise. It is those features that help to decide such matters as when the drawing will commence, that the first number drawn will be an '8', etc. It is those sorts of features that serve to take some people off-track for winning altogether and leave others, like William, on. And it is exactly those features that the agent is able to replicate in a non-stealing alternative. In this way, the agent, should she happen to be so inclined, can *both* not steal *and* boost William's chances of winning.

The best examples of the spatial-temporal-mechanical features that help to decide in the end who wins and that the agent is in a position to grasp in advance are those that the agent both *wants* her yet-to-be-performed actual stealing to have and has the wherewithal to *make* her actual stealing have. For example, she may intend, for purposes of

[35] This is not, I think, just a coincidence, but rather something we recognize intuitively and pre-analytically, and the basis for our thought that William, the slave-child and people who exist and suffer under depletion have been harmed.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

avoiding detection, to 'hurry through' the actual entering of the code into the BlackBerry to transfer the money; and she may have the real ability to do just that. Just as she can insure that her actual stealing will have the feature of being 'hurried', she can equally well, consistent with all applicable epistemic and other constraints, more or less insure that her non-stealing alternative will have that feature, too. She thus can 'hurry through' her act of using her BlackBerry to contribute, from her own personal account, to Save the Children. If it is true that the fact of her actual stealing being 'hurried' will boost somewhat William's chances of winning, then the agent after all has a way of similarly boosting somewhat William's chances of winning and yet not stealing: she can simply select a non-stealing alternative that is itself 'hurried'.

This point has nothing to do with whether the agent is inclined to, or would, or even whether she ought to, choose any such helpful-to-William non-stealing alternative. We thus need not imagine a change of heart on the agent's part; and we are free to suppose that, if she had not stolen half the jackpot, she 'would' have absconded with the whole *and* blown up the building in which the drawing was to have been held. And no one thinks she is obligated to boost the chances that William, as opposed to any other equally deserving purchaser of a lottery ticket, will win the full jackpot. But the expectational account is strictly a test for harm. As a maximizing test, it gets to that issue by canvassing *all* the alternatives that exist for the agent at the critical time and comparing those alternatives to the actually performed act. On that test, only the *existence* of the better alternative for the agent at the critical time matters, and not whether she wants to create, or would have created, or is obligated to create, additional expected well-being. Likewise, I harm you when I shoot you in the arm, because I have the alternative of not shooting you at all, whether or not that alternative is one that I want to or would or ought to do.

The objection is, thus, in effect correct and a caveat is in order: the agent did have certain non-stealing alternatives that would have boosted William's chances of winning above the base level of something like 1 in 10^{1000}. It all depends on *how* she doesn't steal. If she performs any among the 'best' of the non-stealing alternatives – those alternatives, that is, that have the various spatial-temporal-mechanical features that the agent is in a position to anticipate in her own actual though yet-to-be-performed act of stealing – she will thereby succeed in boosting William's chances of winning somewhat beyond that base level.[36]

[36] Contributing to Save the Children instead of stealing is just the sort of 'dummy' act that is better on its face for William but that is the sort of alternative in which the agent is able to replicate the spatial-temporal-mechanical features that she is able to anticipate in her actual act of stealing. So is feigning to steal, or just standing there

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

However, we should also realize that this caveat in point of fact takes us *very little distance* toward establishing that the agent's stealing harms William. Clearly, the agent will be in a position to grasp, prior to performance, *very little* about her yet-to-be-performed act of stealing and the many spatial-temporal-mechanical features that that act will turn out to have and that will help to decide who wins the lottery. In fact, just prior to performance, it may not even be a determinate matter of fact, for many such spatial-temporal-mechanical features, whether the particular act of stealing the agent will in the end perform will have those features or not. For example, suppose that, as things turn out, the agent steals the money using her BlackBerry in an efficient 20.73 seconds. That precise detail is not – except in an unusual case[37] – something she is in a position, prior to performance, to grasp. A stopwatch can tell her *after the fact* that her act endured for 20.73 seconds. But it can't penetrate that particular veil of ignorance *prior to performance*. That feature being obscure to the agent at the critical time, it is not one that she can replicate in a non-stealing alternative.[38]

taking deep breaths. However, should the agent perceive midcourse that the Save the Children contribution seems to be taking longer than she had projected for her stealing, she has the option of aborting that sequence and moving to some other non-stealing, better-for-William act instead, such as just standing there.

Of course, no plausible normative theory will require that the agent not steal in a way that replicates how she would steal if she were going to steal. And there is no reason in the world why the agent might want, or be inclined, to not steal in that particular way. Nor do we have grounds to accept the counterfactual that, if the agent were to have chosen not to steal, she would have not stolen in way that gives some advantage to William's winning. But under an expectational account or any other maximizing account, those points are irrelevant to the question of harm. Harm is decided, rather, on the basis of what alternatives existed for the agent at the critical time. The fact that the agent performed an act that created a certain amount of expected well-being for a given person when she had the alternative of performing an act that would have created more is therefore directly relevant to the question of harm. And it is relevant to the question of harm even if the agent would not have, does not want to and morally need not perform that alternative act.

[37] In the unusual case in which the agent does happen to anticipate that her stealing will take exactly 20.73 seconds – perhaps because she wants it to last exactly that long and has the wherewithal to make it last exactly that long by keeping a close eye on the stopwatch – the points just made will not hold. If the agent knows in advance her stealing will take exactly 20.73 seconds, then she can replicate that feature – using the same stopwatch – in a non-stealing alternative. By doing so, the agent will thereby boost William's chances of winning to the same extent that the fact that the agent's stealing lasts exactly 20.73 seconds boosts William's chances of winning. Nonetheless, we can be sure that, even in the unusual case, the agent's yet-to-be-performed act of stealing will inevitably have many other spatial-temporal-mechanical features that help to decide in the end that William will win but that the agent cannot anticipate – and thus cannot replicate – in a non-stealing alternative.

[38] More precisely, she is in no position to grasp in advance that the particular act of stealing that she will perform, among the many concrete alternatives that fall under that general heading and have the spatial-temporal-mechanical features that she is in a position to anticipate in advance, will be one the duration of which is 20.73 seconds.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

At the same time, it is exactly the kind of feature that may well help to decide who wins in the end.

The import of the caveat is thus highly limited. In the end, the agent will only have the ability to boost William's chances of winning by some very small amount – from 1 in 10^{1000} to something like (estimating wildly now) 1.00001 in 10^{1000} – under the best of the non-stealing alternatives. Despite the caveat, then, things thus remain grim for William. If the agent doesn't steal, William *almost can't* win. How then can her stealing *harm* him?

C. Two bad inferences to a 'no harm done' result

All we have shown so far is how little expected well-being the agent can create for William even under the best of her non-stealing alternatives. Also critical to the harm assessment will be a comparison between that bleak prospect and the prospect that is created for William by the agents' choice to steal. The ultimate claim, of course, of the can't-expect-better argument is 'no harm done'. But what, precisely, is the comparison that supports that result? Two possible answers to this question may seem natural here, and we may, on initial inspection, think that one or the other – or when fully under the spell of the non-identity problem perhaps we simply oscillate between the two – is at play in the can't-expect-better problem.

> *Argument I* ('Can't-expect-better-than-*get*-at-actual-outcome' argument against harm to William):
>
> The agent's stealing has in fact achieved for William half a billion dollars! But the amount of expected well-being that would have been created under the best of the agent's non-stealing alternatives is very low! So: (1) the actual well-being created for William, given the agent's stealing, is greater than the expected well-being created for William, given the best of the agent's non-stealing alternatives. (2) The agent's stealing is thus better for William than is the agent's performing any such non-stealing alternative. (3) The agent's stealing therefore does not harm William.

This argument assesses harm by comparing William's actual well-being, given the agent's stealing, against his expected well-being, given the agent's performance of any one of the best of the non-stealing alternatives. That actual well-being is clearly *greater* than are any of those alternative expected well-being levels. A bird in hand is worth two in the bush. Half a billion is *more* than a very small chance at a full billion. So the agent's stealing is after all *better* for William than

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

are even the best of the agent's non-stealing alternatives. The agent's stealing, therefore, does not *harm* William.

The problem with Argument I is that the comparison it makes use of in order to move from (1) to (2) and (3) is irrelevant to the question of harm – and to the question of when one act is better than another. At least, it is not the comparison contemplated by the expectational account of harm – or, for that matter, the objective account. It rather relies on a third, highly problematic, *hybrid* account that decides harm by reference to a comparison between how things actually turn out for William, given that the agent steals, and the expected well-being created by the best of the agent's non-stealing alternatives. But we should question that account. In the very case at hand, we run into trouble just by adding the 'against all odds' assumption that William would have won the lottery had the agent not stolen. In that case, the actual well-being for William created by the agent's not stealing is quite high. But we have left all the expected values as they were before. Thus, the expected value of the agent's *not* stealing remains quite low, and, we can also note, so does the expected value of the agent's *stealing*. Since the actual value of the agent's not stealing is here greater than the expected value of the agent's stealing, the hybrid account would now seem to imply that the agent's stealing harms William. But we have already derived, applying that same hybrid account, that the agent's stealing does *not* harm William.[39] Further excavation suggests other problems as well. Given just the anti-symmetry of betterness, for example, we cannot say *both* that the agent's stealing is better for William than her not stealing *and* that her not stealing is better for William than her stealing. But the hybrid account would seem to imply both of these results, given the 'against all odds' assumption introduced above. The hybrid account, in effect, allows us *too* much flexibility. In some cases, we can reach any result about harm we want by appealing to such an account – and when we go too far we can reach a contradiction.

[39] It might be argued that, since the act of not stealing – more precisely, the act of proceeding to draw the balls without stealing – is not performed, it has no 'actual value'. But that act does have a perfectly determinate value for William, if it is performed, given the 'against all odds' assumption. And it is that counterfactually established determinate value that is sometimes relied on by those theories that take an actual value approach to betterness and moral obligation generally. For example, in their description of 'actual-outcome value', Oddie and Menzies write that 'the value of an option in the case in which it is realized is the value of the actual outcome the option (usually in conjunction with external chance factors) produces; and in the case in which the option is not realized, its value is the value of the outcome the option would have produced had that option been chosen' (Oddie and Menzies, 'An Objectivist's Guide to Subjective Value', pp. 512–16).

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

A second argument may seem more credible:

> *Argument II* ('Can't-expect-better-than-*expect*-given-actual-choice' argument against harm to William):
>
> It has already been established that William's chances of winning under even the best of the non-stealing alternatives are very low. But – even given that the agent's stealing does not *secure* a win for William (it is no 'bird in hand') – surely the agent's act of stealing makes William's chances of winning greater than *that*! Surely, indeed, the agent's stealing is *necessary* to those chances being as high as they in fact are! And surely that boost to William's chances of winning is significant enough to balance out the fact that William will collect twice as much at any outcome at which he wins and the agent does not steal as he will at any outcome at which he wins and the agent does steal. So: (1) The expected well-being created for William by the agent's stealing is greater than the expected well-being created for William by the best of the non-stealing alternatives. (2) The agent's stealing is thus better for William than is the agent's performance of any such non-stealing alternative. (3) The agent's stealing therefore does not harm William.

Argument II relies on an expected-against-expected comparison – the very sort of comparison that the expectational account explicitly approves. And we have already agreed that the chances of William's winning under a non-stealing alternative – even in the best case, where the agent replicates all the spatial-temporal-mechanical features that she is in a position to grasp prior to performance in her yet-to-be-performed act of stealing – are very low. Surely – the argument proceeds – the agent's stealing makes the chances of William's winning significantly *greater* than *that*. So the agent's stealing must be, expectationally speaking, *better* for William than the agent's performing any such non-stealing alternative. The agent's stealing, therefore, does not *harm* William.

But Argument II is, like Argument I, defective. In Argument II, the problem is with premise (1). Premise (1) presupposes that the agent's yet-to-be-performed act of stealing somehow makes William's chances of winning greater than they are even under the best of the non-stealing alternatives. But where those chances are calculated, as they must be, *at that critical time just prior to performance*, that claim is not one we can accept. After all, whether the agent steals or not, she must draw a *thousand balls precisely in order* if William is to win the lottery. Why aren't the chances that the agent will succeed at that task unbelievably low even if the agent does steal? Why aren't the chances that the agent

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

will succeed at that task, if she steals, no greater than they are under the best of the non-stealing alternatives?

Now, a critic may feel that there is surely *something* about the agent's act being an act of stealing that gives some slight, ineluctable, edge to William's winning. The critic may sense that the act's being a *stealing* is somehow *necessary* to William's having whatever very low chance of winning he has. It is that very feeling that makes the problem with premise (1) hard to detect and the argument itself a fallacy. We thus need to understand and examine this feeling carefully.

Perhaps the thinking is that the act being an act of stealing helped to decide, for example, that the act would be 'hurried', and that the act being 'hurried' was, as noted in section IV.B above, something the agent plausibly was in a position to grasp at that critical time prior to performance. Perhaps the agent grasped that her act would be 'hurried' because, as we noted earlier, she wanted it to be 'hurried' and had the ability to make it be 'hurried'. But that means that it is correct, after all, for the probability calculation to reflect the effect of the act's being 'hurried' on William's chances of winning. Moreover, the feature of being 'hurried', as we noted earlier, is just the kind of spatial-temporal-mechanical feature that may well have helped to decide that in the end William would win. William's chances of winning thus can correctly be deemed – at the critical time – to have been boosted by the act's being 'hurried' – boosted, that is, beyond the base level of 1 in 10^{1000}. On these grounds, the critic might thus conclude that William's chances are, after all, boosted – boosted enough to make premise (1) true.

But such a conclusion would be mistaken. That a *boost* to William's chances of winning is created by the agent's stealing having the feature of being 'hurried' does not show that the act's being an act of stealing makes William's chances of winning *greater than* they are under the best of the agent's non-stealing alternatives. As also noted in section IV.B, certain strategies are available to the agent at the critical time, consistent with her not stealing, that would likewise have boosted William's chances of winning. On examination, those strategies boost William's chances of winning under a non-stealing alternative *just as much* beyond the base level as the act of stealing being 'hurried' itself boosts William's chances of winning. After all, the feature of being 'hurried' is exactly the kind of feature the agent is able to *replicate* in a non-stealing alternative. The agent can, for example, 'hurriedly' transfer funds from her own private account to Save the Children instead of stealing half the jackpot. Of course, if she is not going to steal half the jackpot, then she may have no particular incentive to make a non-stealing transfer of funds, or perform any other alternative, in a manner that is 'hurried'. And certainly we do not think she would be particularly obligated to perform such a 'hurried' alternative just so

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

that she can boost somewhat *William's* chances of winning. But the expectational account of harm is a maximizing account. So it does not matter that the agent *would* not have done things in a 'hurried' way; and there is no claim that she *ought* to have done things in that way. What matters, rather, is that she had the *alternative* of doing things in that way at the critical time.

The upshot is that the fact that the agent's yet-to-be-performed act is an act of stealing does not after all make William's chances of winning any greater than they are under the best of the agent's non-stealing alternatives. More generally, when we compare the chances that William will win, given that the agent steals, against the chances that William will win under the best of the non-stealing alternatives – that is, under those non-stealing alternatives in which the agent has conscientiously replicated each and every spatial-temporal-mechanical feature that she is in a position to grasp in advance in her own yet-to-be-performed act of stealing – we see no grounds to think that the act's being a stealing gives any sort of categorical advantage to William over each and every such non-stealing alternative. If, in light of the boost to William's chances of winning created by the act of stealing being 'hurried', 1.00001×10^{1000} is the right number for that act, then there is no reason to think that that is not the right number – at that critical time just prior to performance –for the best of the agent's non-stealing alternatives as well.

The critic might try to press the point further, making some broad gesture toward the many additional spatial-temporal-mechanical features the agent's yet-to-be-performed act of stealing will surely have beyond those obvious and crude features (like being 'hurried') the agent is in a position to grasp prior to performance. It is surely correct that all those spatial-temporal-mechanical minutiae may well help to decide that William will win the lottery rather than someone else. Should we not then conclude that the agent's act of stealing after all makes William's chances of winning greater than they are under the best of the non-stealing alternatives?

No. The reality is that the agent is able to grasp, prior to performance, *very little* about her yet-to-be-performed act of stealing and the many spatial-temporal-mechanical features that that act will have and that will in the end help to decide who will win the lottery. We noted this point in section IV.B above for the purpose of showing just how little the agent can do under a *non-stealing* alternative to boost William's chances of winning. (The caveat on the claim that William had almost no chances of winning under a non-stealing alternative was, we saw, highly limited.) We now need to note that the same point applies to show just how little the effect is on William's chances of winning when the agent opts for a *stealing* alternative. Prior to performance, many of

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

the spatial-temporal-mechanical features that the yet-to-be-performed act of stealing will have remain completely obscure to the agent, who has no crystal ball. Indeed, that her act will have the specific spatial-temporal-mechanical features that it turns out to have may not even be, at that critical time just prior to performance, a determinate matter of fact. But that means that the fact that the act of stealing will turn out to have those many features must be *screened out* of the calculation of the probability – which calculation must be made as of that critical time just prior to performance – of William's chances of winning under a stealing alternative. That those features will *in the end*, as things turn out, help William to win, in other words, is a fact that cannot be permitted to bear on William's chances of winning *at the beginning* – i.e., at that critical time just *prior to performance*.

Suppose, for example, the agent's stealing turns out to last exactly 20.73 seconds. That the agent's yet-to-be-performed act of stealing will have that feature is nothing that the agent is in a position to grasp prior to performance.[40] But that means that her act of stealing's having that particular feature cannot be counted as boosting, at the critical time just prior to performance, William's chances of winning. That the act of stealing will turn out to have that feature – and many others – may well emerge *after* the act is performed. We, at that juncture, may be able to trace how it happened that the act's being a stealing led to the act's being 'hurried, which in turn led to the act's lasting exactly 20.73 seconds, which in turn helped William to win. But we cannot correctly allow the calculation of how likely the agent's stealing makes William's winning – calculated at that critical time just prior to performance – to be tainted by that highly specific, after-the-fact information.

To put the point another way, the feature of being 'hurried' is more the exception than the rule. If the agent is in a position to grasp that her act of stealing will have a particular feature – such as that it is 'hurried' – then she can replicate that feature in a non-stealing alternative. If she isn't, then that her act of stealing will turn out to have a particular feature is nothing that counts in favor of that act of stealing at the critical time. What, then, of those features – such as 'being an act of stealing' – that the agent can grasp, in her own yet-to-be-performed act of stealing, prior to performance yet cannot replicate in a

[40] Of course, there may be the unusual case in which that assumption does not hold, and the agent is able to grasp in advance, and it is a determinate matter of fact, that her act will last exactly 20.73 seconds. But we have already worked through what to say about that case. If the agent does happen to grasp in advance that her actual stealing will last 20.73 seconds – if, for example, she wants it to last exactly that long and uses a stopwatch to insure that it lasts exactly that long – then her act lasting exactly that long becomes the sort of spatial-temporal-mechanical feature she can replicate in a non-stealing alternative. See n. 37.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

non-stealing alternative? As noted in section IV.B above, those features are not, on inspection, the sorts of features that help to decide that William rather than someone else will win the lottery. It is the spatial-temporal-mechanical features that do that important work. Hence, it does not matter that an act being an act of stealing is not a feature that the agent can replicate in a non-stealing alternative.

The upshot is that Argument II must be rejected as well. The fact that the agent's yet-to-be-performed act is an act of stealing does not after all make William's winning more likely than it is under the best of the non-stealing alternatives. If the agent, in other words, goes about not stealing in a 'best' way, she can create just as much of a chance, calculated at that critical time just prior to performance, for William to win – i.e. precious little – as she can by stealing.

We should thus reject the idea that the expected well-being created for William by the agent's stealing is greater than the expected well-being created for William by the agent's performance of a 'best' non-stealing alternative. In rejecting that idea, we have not even yet begun to exploit, but legitimately could, the fact that the jackpot, if William does happen to win and the agent does not steal, will be twice the size of the jackpot he in fact has won.

These results pave the way for a finding of harm-to-William under an expectational account. This is so even if we include in that account an objective element to reflect the idea that 'misses don't count'. For in the case at hand there is no miss.

* * *

The lottery example thus divides into two beguiling but mistaken arguments. Argument I is the 'can't-expect-better-than-*get*-at-actual-outcome' argument. But that argument relies on a problematic hybrid account of harm – an account that does not reliably indicate either betterness or harm. Argument II is the 'can't-expect-better-than-*expect*-given-actual-choice' argument. The fallacy of this latter argument goes one step deeper. The harm account that it relies on does not itself seem problematic. The problem arises, rather, when we allow information regarding how the future in fact unfolds to taint our calculation of the chances of William's winning, and hence our calculation of expected well-being, created by the agent's stealing. That calculation must be restricted to information the agent is in a position to grasp at that critical time prior to the performance of the act under scrutiny. It is always hard, when we know something, to keep firmly in mind that we must proceed as though we don't. But once we take care to regulate our thinking in that way, it becomes no longer plausible to deny that, among the many non-stealing alternatives that exist for the agent at that

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

critical time just prior to performance, there exists at least one – whether or not it is one she would, wants or ought to perform – that makes it just as likely (i.e. not very) that William will win as the agent's stealing does.

V. THE DEPLETION EXAMPLE

A. Parfit's argument to a 'no harm done' result

In Parfit's example, agents are to choose between two resource policies, one that favors depletion of the earth's resources and one that does not. While a choice of depletion will, if anything, benefit members of the present and the next few generations, 'after one or two centuries' it will also lead to a substantial decline in well-being across wide swaths of the population.[41] For that reason among others, we think (and stipulate here) that the choice of depletion over non-depletion (that is, 'conservation') is wrong. But at just that point at which depletion begins to exact its toll, and as a natural consequence of the fact that 'the choice between our two policies would affect the timing of later conceptions', 'there would be no one living . . . who would have been born whichever policy we chose'.[42] But that means that the very people whose well-being levels are reduced by the depletion choice 'owe their existence' to that choice.[43] Thus, while conservation would have made people better off, the people who are better off would also have been distinct people, 'non-identical' to those who suffer from depletion. At the same time, the lives of the people who do exist are, though burdened, unambiguously 'worth living'.[44] Thus, 'our choice [of depletion] will not be worse [than conservation would have been] for anyone who ever lives'.[45] The depletion choice, in other words, is bad for and harms no one at all. Yet it is still wrong. The person-based intuition must therefore be rejected.

Some of Parfit's language suggests that the depletion example is itself a variation on the won't-do-better, or counterfactual, problem. But a better reading of the argument will not understand it to rely on whether the apparent victims of the depletion choice – that specific collection of people who both exist and suffer under that choice; we can call them the 'D-people' – *would have existed* had a conservation alternative been chosen in place of depletion. Perhaps it is true in this case that, had depletion not been chosen, no D-person would ever

[41] *Reasons and Persons*, pp. 361–3.
[42] *Reasons and Persons*, p. 361.
[43] *Reasons and Persons*, p. 361.
[44] *Reasons and Persons*, p. 362.
[45] *Reasons and Persons*, p. 363.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

have existed. But as we have seen we cannot infer from that fact that no D-person is harmed.[46] Nor, apparently, is the argument intended to rely on an objective account of harm. For that account, as in the lottery example, would reject the inference to the 'no harm done' result. There is nothing in natural law or the acts of other agents that bars the agent in the lottery example from both not stealing and drawing a winning-for-William number. Likewise, there is nothing in natural law or the acts of other agents that bars the agents of depletion from implementing some conservation alternative that would have led – as the future unfolds at least at some worlds accessible to the agents – to the existence of at least some D-person.

The argument is thus best read to rely on the expectational account of harm. Consider a particular D-person – Jaime – who has come into existence many generations after the depletion choice was put into effect. It is surely the case that the shift from depletion to a conservation alternative would have made Jaime's coming into existence highly improbable.[47] And it is surely better for Jaime that he enjoy a flawed existence – or better, at least, to be given a better chance of a flawed existence – than that he be given almost no chance at existence at all. From there we are to infer that the choice of depletion is not bad for and does not harm Jaime. By exactly the same line of reasoning, we are to infer that the choice of depletion is not bad for and does not harm *any* D-person at all.

But just as we have unraveled the inference to the 'no harm done' result in the lottery example, we can also unravel that same inference in the depletion example. I will thus argue that a part of why we remain confident, in both cases, that a wrong has been done is not that we accept an impersonal account of what agents ought to do but rather that we after all detect, in a way that the can't-expect-better problem camouflages and that we must work to clarify, that the depletion choice *harms* future people.

[46] In fact, a counterfactual reading of the depletion problem substantially weakens the force of the argument. After all, whether any particular D-person – and we need show harm only to one such person – 'would have existed' under conservation or not depends on just how agents implement conservation. And there surely existed some way of implementing conservation that 'would have' led to the existence of at least one such D-person. So Parfit's claim would have to be that agents 'would have' implemented conservation in some other way – some way that would not have brought any D-person into existence. But that claim itself is perilously close to the objectionable won't-do-better problem.

[47] Kavka's reference to probability is more explicit than Parfit's. Thus, Kavka notes that it is 'enormously improbable' that agents would have produced the same child as a non-slave 'even if they had tried'. See 'The Paradox of Future Individuals', p. 100 n. 15.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

* * *

The depletion example is a thicket of details. We can, however, use the simpler lottery example as a map through that thicket to a finding of harm. I will suggest that the argument against harm fails in the depletion example just as it does in the lottery example, and that the expectational account implies harm in the depletion example just as it does in the lottery example (section V.B). I then focus on the one detail of that expectational account of the depletion example that I think is likely not to seem obviously correct (section V.C). Finally, I point out the essential differences between how Parfit himself conceives the depletion example and the account of the depletion example that I present here (section V.D).

B. A twist away from harm then back again

In the context of the lottery example, the can't-expect-better problem leads us to understand that, even under the best of the non-stealing alternatives – those alternatives, that is, that have the various spatial-temporal-mechanical features that the agent is in a position to anticipate in her own yet-to-be-performed act of stealing – William's chances of winning the lottery are almost nothing, and thus that the agent's performance of any such alternative, in point of fact, would have created very little expected well-being for William. That understanding represents the original *Kavkaian twist* that the non-identity problem subjects us to when we begin to think about whether a particular morally questionable act truly does harm future persons: that William's winning the lottery is highly 'precarious' in the case where the agent does not steal, and hence that the agent's stealing may have been, after all, the best thing for William. We see the same twist in the depletion example. There, too, we are led to understand that Jaime's chances of coming into existence would have been reduced to almost nothing in the event that the agents had opted against depletion – and hence that depletion, after all, perhaps benefits, rather than harms, Jaime.

Once that twist is made, the can't-expect-better problem proceeds along one of two paths: the 'can't-expect-better-than-*get*-at-actual-outcome' argument (Argument I) and the 'can't-expect-better-than-*expect*-given-actual-choice' argument (Argument II). Argument I urges us to agree that the quantity of actual well-being that is created when the flawed life is secured through depletion, or when half the jackpot is won when the agent steals, is greater than that very low quantity of expected well-being that is created when an alternative is performed that yields the very great chance that one will end up with nothing at all. But we know from the discussion of the lottery example that the actual-against-expected comparison that Argument I appeals to – the

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

hybrid approach to harm – is unreliable. And so we turn instead to Argument II.

Argument II claims that it is surely better to be given some real chance of securing the flawed life or half the jackpot than it is to be given an almost insignificant chance at the better life or the full jackpot. According to Argument II, then, the act under scrutiny – the act of stealing; the act of depletion – harms no one. But then, in the context of the lottery example, we took a hard look at Argument II, and unearthed a *second Kavkaian twist* – really a twist *back* to the understanding that Kavka's point about the precariousness of future events cuts both ways. If William's chances of winning are very low under a non-stealing alternative, they are also, calculated at that critical time just prior to performance, very low *even if the agent steals*. The precisely analogous point – we can now recognize – holds for the depletion example as well. The chances that Jaime himself or any other D-person will ever come into existence are also very low, not just under conservation but *under depletion as well*.

Now, to lay the foundation that opens the door to a finding of harm under an expectational account, the claim must not just be that Jaime's chances of existence under depletion are very low but rather that they are at least as low under depletion as they are under various conservation alternatives.[48] A part of the support for that claim comes in the form of a clear understanding of *just how low* Jaime's chances go under depletion. It is not enough to note that, once the depletion act is performed, many other acts and events have to go a certain way over a century or two if Jaime is ever to come into existence at all. It is also critical to note that the choice of depletion itself may be realized through any one of a vast array of concrete depletion acts. Just as the agent's stealing in the lottery example might, at that critical time prior to performance, last 20.73 seconds, or only 15.9 seconds, or etc., so might whatever particular depletion act the agents end up performing deplete available heating oil supplies by 80.25 percent over a twelve-year period, or only 75.03 percent over a fifteen-and-a-half year period, or etc. Just which depletion act the agents will in the end perform is nothing the agents, who have no crystal ball, are in a position to grasp at that critical time just prior to performance. It may not even be a determinate matter of fact, at that critical time just prior to performance, which depletion act the agents will in the end perform.

[48] We could say here, in place of 'at least as low', 'just about as low', in view of the fact that expected well-being turns not on just the chances of Jaime's coming into existence but also on the actual well-being he has at those outcomes at which he does exist. But I do not think, in the end, that we will need to invoke this technical point since I think we can conclude that Jaime's chances of coming into existence are 'at least as low' under depletion as they are under the best of the conservation alternatives.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

Yes, the fact that one particular depletion act rather than some other is in the end performed may well help to decide that it is Jaime rather than someone else who will come into existence a century or two later. Whether the depletion act in the end depletes at a faster or a slower rate may well bear on whether Jaime exists. But the boost to Jaime's chances of coming into existence that is created by the act's having the particular features that it *eventually* turns out to have is nothing that can be correctly reflected in our assessment of Jaime's chances of coming into existence *as of that critical time just prior to performance*. Jaime's chances of coming into existence thus remain very low, even if the agents opt for depletion.

Thus, we earlier conceded (when faced with the Kavkaian twist against harm) that variations between the depletion act in fact performed and any arbitrary *conservation* alternative may well take Jaime off-track for existence altogether. What the argument just made aims to show (and this is the Kavkaian twist back in the direction of harm) is that variations between the depletion act in the end performed and any arbitrary *depletion* alternative may take Jaime off-track for existence quite effectively as well.

Now, we still have not completed the case for the claim that Jaime's chances of coming into existence are no greater under depletion than they are under a conservation alternative. After all, consistent with the point that *many* features the depletion act will turn out to have cannot be grasped by agents at that critical time prior to performance, it is also the case that *some* features that that act will turn out to have surely can be grasped by the agents in advance. Most plausibly, those features consist of features the agents *want* their depletion act to have and have the ability to *make* their act have. Moreover, some of those features will surely help to decide that Jaime rather than someone else will exist.

We conceded the analogous point in the lottery example, noting there that the agent may be able to grasp, in her yet-to-be-performed act of stealing, various features. But, among those features, the ones that help to decide who in the end wins or who exists – the features that really do serve to boost William's chances of winning or Jaime's chances of coming into existence – are limited to those spatial-temporal-mechanical features that help to decide what balls are plucked in what order and which gametes combine with which. The fact that the agent's act of stealing, for example, will be 'hurried' helps to decide, in the end, who wins the lottery – but not the fact that the act itself is a 'stealing'. Yet we also noted that it is exactly those sorts of spatial-temporal-mechanical features that the agent in the lottery example can replicate in a non-stealing alternative. But that means that any boost that such features give to William's chances of winning the lottery

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

can equally well be achieved by the agent's performance of one of the 'best' of the non-stealing alternatives – one of those alternatives, that is, in which the agent has replicated the various spatial-temporal-mechanical features she is in a position to grasp, at that critical time prior to performance, in her yet-to-be-performed act of stealing.

We can now note precisely the same point for the depletion example. The very spatial-temporal-mechanical features that agents are in a position to grasp, at the critical time prior to performance, in their yet-to-be-performed act of depletion can likewise be replicated in a conservation alternative. But that means that any boost to Jaime's chances of coming into existence that is achieved by depletion can equally well be achieved under the 'best' of the agents' conservation alternatives – that is, under those conservation alternatives that have the various spatial-temporal-mechanical features that the agents are in a position to grasp in their own yet-to-be-performed act of depletion. Jaime's chances of coming into existence, under depletion, are very low, as we have already seen, in any case. What we now see is that, to the extent that depletion boosts Jaime's chances of coming into existence somewhat, the agents after all have a way of *not* depleting that will boost Jaime's chances of coming into existence at least to that same extent.

The idea that the agents can replicate in a conservation alternative those spatial-temporal-mechanical features they are in a position to grasp in advance in their depletion act may seem controversial. We thus look at that idea more closely in section V.C below. For the moment, though, a simple example may be enough to show that it is at least plausible. Suppose that the agents of depletion grasp in advance that their yet-to-be-performed act of depletion will provide some sort of incentive for large numbers of people in – say – the United States to move from the north to the south over the next several decades as a result of severe heating oil shortages. Now, that a particular act will provide people with an incentive to make such a move is itself a spatial-temporal-mechanical feature of sorts: it is a *push*. And it is the fact that Jaime's forebears' families on both sides have been *pushed out* of their distinct locations – a fact that ultimately increased their chances of collecting in some common location – that boosts somewhat Jaime's chances of coming into existence many decades later.

When we think about the depletion choice in that light, it seems clear that depletion offers no particular categorical advantage to Jaime over a conservation alternative. For surely agents can engineer a roughly equivalent push, one that falls well within whatever necessarily imprecise parameters they anticipate for their own yet-to-be-performed depletion alternative, into a conservation alternative. Such a push could take the form of a federal law that increases taxes on heating

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

oil use, thereby providing northerners with an incentive to move south over a period of decades. If the depletion act is anticipated, prior to performance, to provide some still stronger incentive to move north to south, then that tax law could always be supplemented by the planning and construction of attractive, energy-efficient communities in the south that would provide an additional incentive for a move. Depletion is not, in other words, the only alternative agents have of incentivizing people to move north to south – and hence not the only way of boosting somewhat Jaime's chances of coming into existence. A range of conservation alternatives can serve that function as well.

Why, then, couldn't the agents of depletion replicate such a 'push' in a conservation alternative? If it is correct that they could, then it becomes implausible to deny that some conservation alternative existed for the agents at the critical time under which Jaime's coming into existence is at least as likely – that is, not very – as it is under depletion.[49]

We thus conclude that the chances of existence, calculated at the critical time just prior to performance, for Jaime will be just as great under at least some conservation alternatives as they are under depletion. But Jaime will have more well-being at any outcome at which he exists and conservation rather than depletion is performed. The best of the conservation alternatives thus creates more expected well-being for Jaime than depletion does. But that means, on an expectational account, that depletion therefore harms Jaime.

Including an objective element in the expectational account of harm does not change that result. For in this case there is no 'miss' – people do eventually exist and suffer as a result of the depletion choice in ways they don't at worlds that were perfectly accessible to agents at that critical time just prior to performance.

Now, there is no claim here that the agents would have trotted out – had they not opted for, or had they been barred from, depletion – to do all that they might to increase the chances of existence for *Jaime*. Perhaps what they *would* have done, or been *inclined* to do, instead is end civilization as we know it. And even if we think that the agents ought to have kept civilization intact and at the same time pursued conservation, we have no reason to think that they *ought* to have strived to figure out some way to increase *Jaime's* chances of coming into existence. However, the expectational test for harm is itself a *maximizing* test. That means that *all* the alternatives that exist for the agents at the critical time, including those that the agents would be disinclined to perform or would refuse to perform, as well as those the agents are not morally required to perform, are relevant to the

[49] I have elsewhere described a version of this argument. See n. 13 above.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

issue of harm. The bare *existence* of even one conservation alternative that creates more expected well-being than depletion does for even one D-person allows us to infer harm under the expectational account.

* * *

We can sum up. The can't-expect-better problem commences with the idea that the actual well-being of the person who obtains the flawed life under depletion is greater than the expected well-being that is generated when that person is, under conservation, given nothing more than a very great chance of getting nothing at all. There is nothing wrong with the arithmetic in this case: the actual well-being is greater than the expected well-being. The error is in thinking that fact fits in a cogent way into a cogent theory of harm. We avoid the enticing Argument I when we recognize that error. But we must now take care not to fall into yet another bog – the still more enticing Argument II. Here it becomes pertinent that, expectationally speaking, depletion hardly *secures* existence for even a single D-person, for example, for Jaime. One might initially think that depletion at least makes Jaime's chances of existence greater than they are under any conservation alternative. But in the context of the depletion example that more modest thought is wrong, too.[50] Once we set it aside, we are able to discern harm under an expectational account.

C. How to boost Jaime's chances of existence under a conservation alternative

Just as the agent in the lottery example can plausibly grasp, prior to performance, that her yet-to-be-performed act of stealing will be 'hurried', so can the agents in the depletion example plausibly grasp

[50] Argument II fails, as noted earlier, because the premise that the act under scrutiny creates a greater chance of existence for a particular person than any alternative act does is false in the context of the depletion example. In other contexts, however, the analogous premise will be true. It has been reported, for example, that in vitro fertilization (IVF) is associated in rare cases with significant birth defects. See Robertson, 'Procreative Liberty and Harm to Offspring in Assisted Reproduction', pp. 9–10. If in such a case (1) the particular use of IVF makes conception more likely (looking ahead from that critical time just prior to performance) than it is under any facially better alternative that then exists for the agents, thereby making the existence of the new person into whom the zygote eventually develops more likely than it is under any alternative, and (2) the anticipated defect is not generally so severe that the overall, life time level of well-being for that person will fall into the negative range, then, under an expectational account, the use of IVF will not *harm* that new person even if the risk of defect eventuates. This analysis supposes that the case is one in which it is the IVF itself that causes the defect and not an inherent genetic or chromosomal defect that the application of IVF is less likely than nature to rule out. If, in contrast, the case is one in which the defect is an inherent genetic or chromosomal defect, then the case should be treated as a can't-do-better problem, a type of problem discussed briefly in sect. VI below.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

in their yet-to-be-performed depletion act certain features as well. Moreover, some of those same features may well help to decide a century or two later who will come into existence, just as the act's being 'hurried' in the lottery example may well help to decide who will win the lottery.

As we have already noted, however, from these facts it does not follow that depletion makes Jaime's chances of coming into existence any greater than they are under the best of the conservation alternatives. After all, the very spatial-temporal-mechanical features that may boost somewhat Jaime's chances of coming into existence can be replicated by those same agents in a conservation alternative to the extent agents are able to grasp them in advance in their yet-to-be-performed act of depletion. It is not the fact that the agents' act is an act of *depletion*, in other words, that helps to decide that Jaime will exist. It is rather the fact that it is a *push*: something that makes people move from one city or state or country to another and that leads eventually to the particular gametes combining that in fact combine.

The goal of this present part is to show that there is every reason to think that agents were in a position to replicate a *like push* in a conservation alternative. It is useful, in this connection, to think about the specific details regarding just how depletion itself managed the feat of bringing Jaime into existence. Let's suppose that agents are in a position to anticipate that their yet-to-be-performed depletion act will constitute a 'substantial' depletion of heating oil reserves. They are able, let's suppose, to anticipate that fact about their act given the extent of the gains they aim to produce, and their ability to reap those gains, for the benefit of their own generation. They can similarly anticipate that the effect on future generations will itself be substantial and – let's suppose – that a 'weighty incentive' will then exist for large numbers of people living in the north to move to the south. The push, in other words, that the depletion act will ultimately create will be a *hard* one.

This much the agents are able (by hypothesis) to grasp in advance. What happens then? Suppose that the act of depletion in fact performed turns out to be an act that severely depletes heating oil reserves by exactly 75.03 percent over a fifteen-and-a-half year period, which fact – when put together with many other details regarding how the future otherwise unfolds – motivates a precise number of families to move away from Maine a precise number of years later, and a precise number of families to move away from Minnesota a precise number of years later. Some of those families happen to end up in – say – Baton Rouge, where energy costs are not so high and people can both eat and stay warm. It is there that Jaime's own forebears – one, let's suppose, with family from Maine and one with family from Minnesota – eventually

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

meet at Baton Rouge High School, marry, have children, grandchildren and, eventually, Jaime.

Now, the particularized sequence of acts and events just described includes many details about how the future unfolds – including details about how the act of depletion itself unfolds. And the fact that the future turns out to include this particularized sequence of acts and events rather than some other clearly helped to decide that Jaime rather than someone else would in the end exist. Yet that the future would unfold in just the way it has – including that depletion would be implemented in precisely the way that it was – is nothing that the agents were in a position to grasp at the critical time just prior to performance. Indeed, at that critical time, that the future would include that particularized sequence of acts and events may well have been a quite indeterminate matter of fact. That in turn means that the fact that the future would unfold in just the way it has cannot, at the critical time just prior to performance, be counted as boosting in any way Jaime's chances of coming into existence.

A way, then, of getting a clear picture of just how likely it is that Jaime will exist, given depletion, is to ask just how likely it is, at that critical time just prior to performance, that this particularized sequence of acts and events – a sequence that begins just as the act of depletion itself commences, that includes heating oil reserves being depleted by precisely 75.03 percent over a fifteen-and-a-half year period, that draws Jaime's forebears' families to exactly the same small town and that ends with Jaime's existence – will in fact unfold, given a 'substantial' act of depletion that itself creates a 'weighty incentive' for future persons to move from the north to the south. The answer is: not very.

Now let's turn to the conservation side. What we want to do here is consider an alternative possible future, or world – but one perfectly accessible to agents at the critical time – that itself includes a particularized sequence of acts and events, one that begins with an act that is consistent with a conservation policy and nonetheless ends, like the original, with Jaime's coming into existence. Consider, then, a world at which the initial act within the sequence is a new heating oil tax that motivates a precise number of families to move away from Maine a precise number of years later, and a precise number of families to move from Minnesota a precise number of years later. Some of those families happen to end up in Amarillo, where people have come to work in an alternative energy industry (harvesting, say, electricity from windmills) and where they may live in cozy and well-swept dug-outs (cool in summer, warm in winter). There Jaime's own forebears eventually meet, now at Amarillo High instead of Baton Rouge, marry, have children, grandchildren and, eventually, Jaime.

Clearly, the particularized sequence of acts and events just described includes many details – details that the agents, prior to the performance

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

of the conservation act that commences the sequence, are in no position to grasp. It is not, in other words, a sequence that is, at the critical time, even remotely likely. For our purposes here, however, the important fact about this sequence is that, however unlikely it may be, it is a sequence that is no less likely, at that critical time prior to performance, to unfold, given a 'substantial' act of *conservation* that itself creates a 'weighty incentive' for future persons to move from the north to the south, than is the sequence that in fact unfolds, given a 'substantial' act of *depletion* that itself creates a 'weighty incentive' for future persons to move from the north to the south. As part of both sequences, agents have created a 'weighty incentive' for people to move north to south. In both cases, agents have engineered a *hard push* – hard enough to move northerners out of Maine and Minnesota and into some warmer place. If the agents of depletion consist, for example, of legislators acting over a period of years, then, consistent with a conservation policy, those same agents might instead have put into a place the new heating oil tax that we have described. If the push the agents are able – very roughly – to anticipate in their yet-to-be-performed act of depletion is a particularly hard one, then they can combine that tax with various alternative energy projects that have their own power to draw northerners to move south. It would be no excuse for those legislators to make the argument that, had they opted for conservation over depletion, they would have been voted out of office by a public disinclined to self-sacrifice. Even if that prediction were accurate, it does not follow that a conservation alternative does not *exist* for those same legislators. More generally, whether depletion is better for the legislators and their contemporaries is immaterial to whether that same choice eventually harms the then-future person Jaime.

Once we take care to assess probabilities in the forward-looking way that we must if facts about harm are to have the action-guiding function that is the whole point of the expectational account, it becomes very hard, I think, to find any support for the claim that Jaime's coming into existence is somehow less likely under the best of the conservation alternatives than it is under depletion. At that critical time prior to performance, the story that we told about Jaime, his forebears and their families' move to Baton Rouge was nothing more than a highly detailed and highly improbable sequence of acts and events. But not one bit more improbable is the story that we could tell about Jaime, his forebears and their families' move to Amarillo. Of course, even under the best of the conservation alternatives, it remains unbelievably unlikely that Jaime will ever come into existence. But, at that critical time prior to performance, it remains just as unbelievably unlikely that Jaime will come into existence under the depletion choice as well.

The depletion choice – really a vast, protracted, complex of acts wrought by numerous agents over many years – is admittedly more

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

complicated than is the agent's act of stealing in the lottery example. But that complexity, on inspection, cuts both ways. The very complexity of the conservation alternatives means that Jaime's coming into existence is very unlikely under even the best of those alternatives. But the depletion alternatives are every bit as complex, and Jaime's coming into existence remains very unlikely, at the critical time, under each and every one of them as well.[51] The best among the conservation alternatives thus create just as much of a chance for Jaime to exist as depletion does.[52]

D. Fitting the results of the expectational account into Parfit's same people/different people distinction between choices

Parfit's depletion example famously involves future persons, while the lottery example's potential victims all exist at the time the act under scrutiny is performed. But under the expectational account of that example, all persons who do or will exist at a world are, in one sense, treated equally: just as *present* people can be harmed, so can *future*

[51] If, in that sense, the depletion act is after all necessary to Jaime's extremely poor chances of existence being as high as they in fact are, then the only way to escape the 'no harm done' result under the expectational account would be to rely on the fact that the actual well-being created by conservation for Jaime at a world where Jaime exists will be greater than the actual well-being created by depletion for Jaime. While that fact is one we can legitimately exploit, it is not, or so I argue, at all obvious that we will need to do so here.

[52] We can give an analogous account of harm in Kavka's slave-child case. In that case, as well, we suppose the 'bad' act to have taken place, and the couple to have brought a child, say, Anna, into existence as a slave. Clearly, the couple there had the alternative of feigning to enter into the slave-child contract and then proceeding to conceive a child, building into the full sequence all those spatial-temporal-mechanical features that they can grasp in advance about their future act of entering into the contract and producing the child. If they anticipate their actual signing to be quick, they can quickly feign signing as well. If they anticipate they will then repair to a nearby hotel to conceive a child, they can likewise perform a 'non-signing' alternative under which they then repair to the same nearby hotel to conceive a child. It is implausible to think that, as of that critical time prior to performance, the feigned 'entering into the contract' that replicates, as nearly as agents are able, the actual entering into the contract would have made Anna's chances of existence any less than they are under a bona fide 'entering into the contract'. This is in part a function of the fact that how the agents go about entering into the contract itself may unfold in one way rather than another; the entering into the contract, as of the critical time prior to performance, may be implemented by the agents through the performance of any number of concrete acts that equally fall under that heading. Exactly which they shall perform is not something they can anticipate in advance. Moreover, it may not even be a determinate matter of fact. The upshot is that the entering into the contract is *not* necessary to achieving that very low chance of coming into existence that Anna has at the critical time. Other, better ways exist that give her that same very low chance. Since existing as a non-slave is better than existing as a slave, the agents can do more for Anna than they do when they enter the slave-child contract. They have therefore harmed Anna when they do less. As in the depletion example, including an objective element in the expectational account of harm does not change those results since Anna's well-being, given that she is a slave, has not been maximized.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

people be harmed – in the comparative, 'worse off', sense of harm that we intuitively think is important in both moral theory and the law.[53] It is this maneuver that Parfit's own account of the depletion example rules out. For him, the depletion example is an intransigently 'different people' case: none of the people who exist under the depletion choice are identical to any of the people who exist under the conservation choice, and so the depletion choice cannot be said to be *bad for*, or to *harm*, any of those people in any ordinary sense. My argument has been that these apparently 'different people' cases can, and should, be understood as 'same people' cases.[54]

Two points have been important to making that way of looking at the depletion example a plausible one. The first is that the expectational account itself is a *maximizing* account. That means that the relevant alternative acts we must scan to determine whether an act harms any

[53] As noted earlier, the expectational account plausibly excludes merely possible people – those who do not and never will exist – from harm. Maybe it is cogent to say that merely possible people can be harmed – as when, for example, they are left out of existence. But if so the harm that is done to them in that sense would not be deemed morally relevant under any sort of person-based approach. See sect. III.C above.

Claims of harm, under the expectational account, are relativized to worlds, where, for any given existing or future person, it is a determinate matter of fact whether that person exists, at least eventually, or not. Resolving the can't-expect-better problem, then, is a matter of coming to a correct calculation that agents can just as easily (not very) achieve a world where Jaime exists and has more well-being as a result of conservation than they can achieve a world where Jaime exists and has less well-being as a result of depletion.

[54] The three types of non-identity problem I describe in this article do not exhaust all of the arguments that have been associated one way or another with the non-identity problem. A non-identity-like analysis is sometimes used, for example, to construct an argument that an apparently harmful act performed *after* the individual whose plight we are concerned with has *already* come into existence does not harm that individual. Such arguments, which involve *explicitly* 'same-people' choices, often arise, for example, in the context of discussions of the ethics of meat-eating. Has the pig who is humanely butchered at the end of a couple of years of fattening by a kind and caring farmer been *harmed* by having been butchered? Of course. In this context there is no argument to be made that the butchering of the pig was itself something that benefited the pig or avoided harming the pig by virtue of the fact that it secures existence for the pig or at least confers a substantial chance of existence on the pig. The pig has *already* come into existence, and refraining from butchering the pig is not going to have any causal influence over that fact. Now, what may make a causal difference to the pig's coming into existence is the farmer's *plan* or *program* of bringing one or more pigs into existence, caring for them and ultimately butchering them for meat that can then be consumed or sold. Such a plan or program, since it is adopted prior to the pig's existence, would be fair game for the non-identity problem. For then we do have a case in which a seemingly harmful act, because it plays a role in bringing the pig into existence, can be argued, under a non-identity analysis, not to impose harm on the pig. At that juncture, however, we also have a version of the can't-expect-better problem – a case with respect to which, I have argued, we can on closer analysis discern harm. But at least in the case of the plan or program the argument against harm itself warrants the title 'non-identity problem' – a compelling problem we have to work to unravel. Not so in the case of the simple butchering of the pig. Even those theorists who otherwise uncritically accept any and all variations on the non-identity problem and just about any tenable account of harm

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

person are not just two (what the agents have done and what they would have done had they not performed the act under scrutiny) but a much wider array. By taking into account *all* the agents' alternatives, in other words, and not just some much narrower set, we are able to see just how particular persons have been harmed, in an intuitive, comparative sense, by the depletion act.

Second, we must be aware of how easy it is to slip into error when putting probabilities to work. Thus, comparisons between actual and expected well-being must be recognized as unreliable indications of harm. And, in calculating expected well-being, we must make sure to jettison after-the-fact information agents may be able to obtain regarding linkages between the act under scrutiny and things having turned out in the way that they did. We can then recognize that the very low chance of a given person's coming into existence being as high as it in fact is – at the critical time just prior to performance – is after all independent, under an expectational account, of the bad act's having been performed, since agents after all have other, better, alternatives for creating at least *that* great a chance of that person's coming into existence.

VI. CONCLUSION

I have argued that the can't-*expect*-better problem does not, after all, support the claim that some 'bad' acts are 'bad for', and harm, no one. That work, however, has no implications for the can't-*do*-better problem, which raises the question of whether it is permissible to bring a child into existence whose well-being, though clearly positive, is unavoidably capped at a low level in a case where (for example) the well-being of each existing and future person, including the child, has been maximized. As suggested in section I, it is not clear to me that that procreative choice would be wrong. It is thus not clear to me that

(even a counterfactual account) should concede that that latter, after-the-fact act clearly counts as one that harms the pig.

Whether the human clones Kazuo Ishiguro imagines in his *Never Let Me Go* are harmed can be analyzed in precisely the same way. The clones begin their lives at Hailsham, where they are kept safe, well-fed and segregated in the countryside away from mainstream society, all so that they can better serve society as organ donors. Once the clones mature, organs are taken from them once, twice, thrice or until they 'complete', that is, die, after living about half a normal life span, or are otherwise "switched off" when important organs have been taken but death does not come. Are they *harmed* when the surgeon takes that first kidney from their bodies? Of course – and no interesting variation on the non-identity argument says otherwise. Does the plan or program, developed prior to their existence and forming part of the causal sequence that leads eventually to their existence, harm the clones? In answering this question, we do face a legitimate non-identity problem. But here again the type of problem we face is the 'can't-expect-better' problem – which means that we can after all, on close analysis, discern harm.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press

the can't-do-better problem is a genuine *counterexample* against the person-based intuition. But many theorists – including Singer, Parfit and others – suggest that it is.[55] Nothing I have argued in this article decides that issue.

The discussion of this article does suggest, however, that we need to be more Humean in our thinking about the bad acts that agents perform and the subsequent suffering that takes place as a consequence of those bad acts. Thus, the stealing in the lottery example, in a sense, leads to William's number being drawn; and the depletion choice leads to the existence of Jaime. But those bad acts are not like drugs that have 'medicinal' properties that endow them with some inherent power to induce a felicitous fate to befall a particular person. More precisely: the extraordinarily pitiful powers that those bad acts do have, when it comes to William's winning the lottery or Jaime's coming into existence, are held in equally pitiful supply by various alternative choices that are, at least on their face, substantially better choices. And that means, in turn, that the agents, after all, could have done better for their apparent victims than they have and thus have harmed them.

The can't-expect-better problem is thus best understood as a *probability* problem, and indeed as a fallacy. As such, it raises no serious questions about how it is that we can harm people whom we by the same act cause to exist. If my analysis is correct, then the can't-expect-better problem can take its place as another in a long line of riveting probability problems that we can in the end mercifully set aside – a result that would in no way diminish its significance in helping us understand the structure of moral theory but that may leave us free to retain the person-based intuition as a basic part of that structure.[56]

robertsm@tcnj.edu

[55] Parfit's 'two medical programmes' example has been especially influential in this case. *Reasons and Persons*, p. 367. Peter Singer discusses what he calls the 'prior existence view', tentatively discarding it in favor of an impersonal approach. See, e.g., *Practical Ethics*, pp. 185–91. See also Buchanan, et al., *From Choice to Chance*, pp. 247–57.

[56] I am very grateful to Fred Feldman, Alan McMichael, Peter Vallentyne and David Wasserman, and to anonymous referees for this and another journal, for their extremely insightful comments on earlier versions of this article.

https://doi.org/10.1017/S0953820807002609 Published online by Cambridge University Press