

JOURNAL *of* ETHICS
& SOCIAL PHILOSOPHY

VOLUME XV · NUMBER 3

August 2019

ARTICLES

- 197 The Invisible Hand from the Grave
Barry Lam
- 222 Revisiting the Argument from Action Guidance
Philip Fox
- 255 Basically Deserved Blame and Its Value
Michael McKenna

DISCUSSIONS

- 283 Against Jeffrey Howard on Entrapment
Jonathan Stanhope
- 291 Dietz on Group-Based Reasons
Magnus Jedenheim Edling

The *Journal of Ethics and Social Philosophy* (ISSN 1559-3061) is a peer-reviewed online journal in moral, social, political, and legal philosophy. The journal is founded on the principle of publisher-funded open access. There are no publication fees for authors, and public access to articles is free of charge and is available to all readers under the CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 license. Funding for the journal has been made possible through the generous commitment of the Gould School of Law and the Dornsife College of Letters, Arts, and Sciences at the University of Southern California.

The *Journal of Ethics and Social Philosophy* aspires to be the leading venue for the best new work in the fields that it covers, and it is governed by a correspondingly high editorial standard. The journal welcomes submissions of articles in any of these and related fields of research. The journal is interested in work in the history of ethics that bears directly on topics of contemporary interest, but does not consider articles of purely historical interest. It is the view of the associate editors that the journal's high standard does not preclude publishing work that is critical in nature, provided that it is constructive, well-argued, current, and of sufficiently general interest.

Editor

Mark Schroeder

Associate Editors

Dale Dorsey

James Dreier

Julia Driver

Anca Gheaus

Tristram McPherson

Colleen Murphy

Hille Paakkunainen

Discussion Notes Editor

Kimberley Brownlee

Editorial Board

Elizabeth Anderson	Philip Pettit
David Brink	Gerald Postema
John Broome	Joseph Raz
Joshua Cohen	Henry Richardson
Jonathan Dancy	Thomas M. Scanlon
John Finnis	Tamar Schapiro
John Gardner	David Schmidtz
Leslie Green	Russ Shafer-Landau
Karen Jones	Tommie Shelby
Frances Kamm	Sarah Stroud
Will Kymlicka	Valerie Tiberius
Matthew Liao	Peter Vallentyne
Kasper Lippert-Rasmussen	Gary Watson
Elinor Mason	Kit Wellman
Stephen Perry	Susan Wolf

Managing Editor

David Clark

Copyeditor

Susan Wampler

Typesetting

Matthew Silverstein

THE INVISIBLE HAND FROM THE GRAVE

Barry Lam

IMAGINE WHAT A COUNTRY would be like if the dead could legally vote. We could pass a law permitting citizens to write a “perpetual vote” into their wills, allowing them to plan a vote for their preferred candidate type, party, or issue, in every election forever, and the government would see to it that these votes would compete with the votes of the living.

We have overwhelmingly good moral reasons to reject such a society. The dead, after all, are exempt from the benefits and harms of current political institutions. They are also perpetually underinformed. They cannot learn about the relevant facts that bear on sound political judgment, such as facts about technological, ecological, demographic, moral, and social changes. To give the dead perpetual political power would dilute the power of the living to shape the institutions that they and their descendants must live with. It could subject future generations to political institutions that operate completely contrary to their needs, values, and choices.

The political philosophers most influential to the founding of the United States recognized these moral considerations. Thomas Paine claimed that rule by the dead “is the most ridiculous and insolent of all tyrannies.”¹ Thomas Jefferson wrote in 1824, “Can one generation bind another, and all others, in succession forever? I think not. The Creator has made the earth for the living, not the dead. Rights and powers can only belong to persons, not to things, not to mere matter, unendowed with will.”²

If we agree with these sentiments that deny political power to the dead, do we have similarly good moral reasons to deny *economic power* to the dead? We certainly do not act like it. Trillions of dollars in the US economy and many legal institutions at all levels are tied up in executing the wishes of the dead in the United States and other common law countries like Britain, Ireland, and Australia. One simple example is the *conditional bequest*. You may require, as a condition of inheritance, that your grandchildren marry within a religious faith, that

1 Paine, *Rights of Man*, 11.

2 Jefferson to Major John Cartwright, June 5, 1824, 48.

your wife smoke at least five cigarettes a day, or that a school be named after you in perpetuity, forbidding a change in name even if the school would otherwise go bankrupt.³ Conditional bequests have a long and entertaining history, and sometimes the conditions remain in effect many decades beyond the testator's death.⁴ The state enforces such conditional bequests on behalf of the dead even when the state has no independent interest in whether the conditions are satisfied, and all affected living parties prefer that the conditions not be upheld.

US law also recognizes a financial instrument called a *dynasty trust*, which allows individuals to secure wealth in a tax-sheltered trust for their descendants, in perpetuity if they choose. Such wealth can grow tax free, is transferred tax free, and can even be shielded from all future creditors.⁵ These dynasty trusts make it possible for heirs of the super-rich to sustain affluent lifestyles that are protected not only from the exhausting need to work, but also from the financial consequences of any poor decisions they might make. At the same time, these individuals do not have full property rights over the wealth in the trust. They could not, for instance, disinherit subsequent generations for any reason, as those rights are constrained by the wishes of the original dead founder of the trust. The irrevocable powers of the original dead founder include the amount distributed to each descendant, who counts as a descendant, what contingencies can lead to disinheritance, and how long the trust is to last. The compounding growth of the assets in such trusts can result in more and more wealth from future generations tied to the wishes of the dead.

A third legal instrument is the *charitable trust*, where the dead can earmark current and future wealth for some purpose considered "charitable."⁶ This term is interpreted generously, and has been taken to include the care of abandoned guinea pigs and the preservation of Huey military aircraft, among other caus-

- 3 See *In Re Estate of Feinberg*, 919 N.E.2d 888 (Ill. 2009), and "Widow Fumes at Order to Start Smoking," *San Antonio Express-News*. In the case of Paul Smith College in New York, billionaire philanthropist Joan Weill offered a \$20 million donation to rescue the school from insolvency on the condition that the school be renamed "Paul Smith-Joan Weill College," contrary to the conditional bequest of Paul Smith, who, decades prior, bequeathed the land on which the college sits. The courts found in favor of the original bequest.
- 4 One of the most famous cases involves Wellington R. Burt, a nineteenth-century railroad industrialist, who famously and spitefully required the state to hold his wealth until twenty-one years after the death of his last grandchild who lived during his lifetime. It took ninety-one years for any of his descendants to inherit his wealth.
- 5 Sitkoff and Horowitz, "Unconstitutional Perpetual Trusts."
- 6 For a compelling recent discussion of the plutocratic tendencies of charitable foundations established by the currently living, see Reich, *Just Giving*.

es.⁷ Charitable trusts, like dynasty trusts, can also earn money on behalf of their dead founders. That is, money earned from returns on trust assets such as capital gains or dividends are placed under the power of the wishes of the dead founder, as executed by trustees and courts. Like dynasty trusts, charitable trusts can also exist in perpetuity, so that trust assets are subject to the will of the dead founder forever. Nonprofit institutions like hospitals, museums, and universities can also have large amounts of their endowments constrained by the wishes of long-dead testators. With the wealth and power of a charitable trust, individuals in the past have decreed things like this: that a new park be built for the enjoyment of all and only white people, that a certain wing of a college dormitory be set aside for housing individuals of Confederate ancestry, and that there be an endowed professorship for the study of parapsychology.⁸

Many political institutions contribute to the enforcement of such dead-hand control. County district attorneys and state attorneys general have staff dedicated to enforcing charitable deeds, conditional bequests, and dynasty trusts. But most dead-hand control of wealth is subject to the rulings of courts. Large institutions, like Princeton University, Paul Smith's College, Newcomb College, Girard College, and the Hershey Company, to name a few, have faced major lawsuits over alleged violations of original donor intent.⁹ In almost all of these cases, with some exceptions, the job of the court is to determine what practice is most consistent with the wishes of the dead as stated in a particular deed or document, not whether the current state of the world is such that the money is better spent according to, rather than contrary to, the wishes of a dead donor.

It strikes me as odd that we are perfectly happy transferring economic power from the living to the dead via these three instruments—as though the reasons for preventing dead-hand control in politics do not carry over to personal wealth. Even more puzzling is that we prevent dead-hand control in other sectors of the economy. Founders or owners of businesses cannot stipulate legally binding *business* decisions to be carried out after their death. For instance, a founder cannot posthumously require that manufacturing be conducted in her hometown, or that the company will never do business in Japan. Such founders are perfectly within their rights to make such decisions while they are alive, but there are no

7 Cambridge Cavy Trust is one example of a guinea pig trust. The Huey aircraft example is from Madoff, *Immortality and the Law*.

8 See, for example, *Evans v. Newton*, 382 U.S. 296 (1966); *Evans v. Abney*, 396 U.S. 435 (1970); and *Tennessee Div. of the United Daughters of the Confederacy v. Vanderbilt Univ.*, No. M2003-02632-COA-R3-CV (Tenn. Ct. App. May 3, 2005).

9 Girard College is the result of the most "litigated will in history." Interestingly, the Hershey case, also in Pennsylvania, does not trail far behind.

legal institutions set up to execute a businessperson's wishes postmortem, and certainly no way for them to execute such wishes in perpetuity. As in the case of political rights, there is no particular mystery as to why. Markets and fashions change, sometimes drastically, and failing to change business practices out of some sense of loyalty or obligation to a past owner can lead very quickly to the death of the business. Requiring the state to enforce a founder's preferences can be tantamount to requiring the state to lead a business to its death.

It is a curious form of greed to feel entitled to eternal ownership of anything, and even more curious that the living permit and even promote it. To say the least, the idea that personal wealth may be tied to the wishes of the dead in perpetuity is in need of justification.

The philosophical literature on this topic is equivocal. Some philosophers believe that individuals can suffer posthumous harm, can be victims of posthumous moral wrongs, and are thereby afforded posthumous rights.¹⁰ These views, if they succeed, can be one way to underwrite the legitimacy of state enforcement of posthumous property rights.

On the other hand we have the considerations of Thomas Paine and Thomas Jefferson that fully generalize to economic power: the world belongs to the living. On this view, death is the final and ultimate alienation of all rights, including rights to personal property. As suggested by the cases of politics and private business, ideas about what it is good to do with power and resources ought to change with the changing conditions of the world. When it is in no living person's best interest for a sliver of the economy to be dedicated to the study of parapsychology, there should be no legal mechanism compelling expenditure on an endowed chair in this field. In general, the fact that a dead person willed that some pot of money be spent in a certain way is never a sufficient justification that the money be spent in that way.

These considerations are far from theoretical. As wealth inequality increases, and as returns on investment far outpace productivity growth, the wealthy are earmarking ever-larger amounts of money from the future economy to carry out their current wishes.¹¹ The result could be a future economy that reflects the preferences of a past aristocracy rather than the majority of those living. Philosophers, even libertarians, have long worried about the threat that wealth inequality poses to justice when it can be perpetually inherited rather than earned.¹² Inheritance is certainly a concern. But there is a more general concern I have that

10 Feinberg, *The Moral Limits of the Criminal Law*; Pitcher, "The Misfortunes of the Dead"; Levenbook, "Harming Someone after His Death"; Luper, "Posthumous Harm."

11 Masterfully documented by Picketty, *Capital in the Twenty-First Century*.

12 Otsuka, *Libertarianism without Inequality*.

reaches beyond the mere persistence of wealth inequality through inheritance. The concern is that there is a deep injustice inherent in disenfranchising the living by redistributing control over wealth to the dead, leading to the tyranny of the *past* opulent over the majority.

Ultimately, I contend that there is no justification for posthumous rights. The extent to which we feel that we owe anything to the dead can be explained completely in terms of the best interests of the living; to the extent that the wishes of the dead depart from the wishes of the living, there is overriding reason to reject any felt need to honor them. I believe that all of the reasons we have to reject the institutional handover of power from the living to the dead in politics and private business carry over to all power in all areas. The legal institutions we have erected to carry out the wishes of the dead currently in the common law countries are therefore unjust and must be reformed.

1. FREEDOM OF TESTATION AND PERPETUAL CONTROL

To understand how we have arrived at our current legal practices of wealth transfer, we need to grasp two crucial concepts. The first is *freedom of testation* and the second is *perpetual control*. Freedom of testation is the idea that your right to property includes the right to determine what happens to that property after your death. Perpetual control is the idea that the right to determine what happens to your property after death extends forever into the future. I argue that neither perpetual control nor freedom of testation are defensible as principles on which to base a legal framework for the disposition of private property after death.

1.1. *A Brief History of Perpetual Control and Its Problems*

If you have any interest in English literature, you will be familiar with the following plotline. A protagonist, be it a smart young woman or able young man, is unable to secure part of their father's estate due to some antiquated inheritance line established decades or centuries ago. Instead, some evil, sociopathic, or undeserving relative is the inheritor, requiring our protagonist to journey through life using only their entrepreneurial spirit or personal charms to secure their own living, resulting in a tale of the triumph of rugged individualism, or the tragedies of class stratification. This generic plot has as its setting a world in which the living are burdened by something called a *fee tail* (or *entailed estate*). The fee tail was an instrument by which a wealthy landowning individual could, through deed or custom, set up a perpetual, state-enforced inheritance line for an estate. On the one hand, subsequent inheritors were legally entitled to the estate as long as

they were stipulated to be in the line of succession in the deed, *regardless of what an inheritor from a previous generation had done with the estate*. On the other hand, no subsequent inheritor had full property rights to the estate. They could not, for instance, decide to divide it evenly between disinherited siblings or revise the line of succession. No matter what subsequent inheritors wanted or needed, the original landowning individual's wishes would continue to be executed in subsequent generations.

Initially, fee tails traced back to feudal traditions rather than an act of will by an individual landowner. The fee tail was one of the last remnants of feudalism in Europe, with lines of succession more or less determined by customs and history, and almost always along patriarchal lines. Then, in 1540, Britain passed the Statute of Wills, giving individuals a degree of freedom of testation.¹³ This statute enabled landowners to draft fee tails *by decision*. Lines of succession would be chosen and codified into law from wills.

The Statute of Wills and the fee tail were in fact strange bedfellows. The statute represented a move away from feudalism and toward a market economy in Britain, where more robust individual property rights and the liberalization of real estate markets were the primary concerns of newer members of the landed gentry. If a piece of land was held in a fee tail, it was not marketable to purchasers. Such property was either completely unalienable (i.e., not saleable at all and available for ownership only by those in the line of succession), or saleable but subject to state intervention (that is, seizure—the state would need to “return” a purchased piece of land to a member of the line upon the death of a previous inheritor). Another way of seeing it is that land could not be easily converted to cash. Newer landed gentry wanted access to entailed land, and they wanted full rights of ownership over the land they acquired, but fee tails prevented such rights. The fee tail was in many ways completely contrary to the capitalist ethos emerging in Britain at the time.

On the other hand, once new landowners had acquired land, and had full rights of testation over the land, new landowners often wanted to become feudal landowners in their own right, keeping land within the family in perpetuity. What they found objectionable in fee tails as buyers they found perfectly attractive as owners. The Statute of Wills simultaneously expressed the right of landowners to control their property after their deaths, contrary to strict feudal inheritance traditions, and allowed new landowners to control their property completely in accordance with those feudal traditions if they chose.

Against this backdrop, the *Rule against Perpetuities* arose in the late seven-

13 According to the statute, two-thirds of one's estate could be freely testated. Baker, *The Oxford History of the Laws of England*.

teenth century, allowing fee tails in principle but curtailing them from extending too far into the future. In a famous case involving the Duke of Norfolk, a tribunal of judges executed a rule placing a time constraint on an individual's right to property after his death, a precedent that has survived over three hundred years.¹⁴ In the common law, the eventual formulation of "lives in being + 21 years" was the limit to the length of state enforcement of the wishes of the dead in fee tails. Very roughly, this means that the state will enforce the wishes of the dead for no longer than the length of the lives of the youngest people affected by the deed or trust, plus twenty-one years, which for practical purposes is about one hundred years.

By the time authors such as Jane Austen and George Eliot wrote some of our favorite English novels against the backdrop of the entailed estate, fee tails were mostly subject to the Rule against Perpetuities. Nonetheless, as such novels illustrate, the social harms of such instruments persisted from the times of feudalism. Some of the damaging effects of fee tails and feudalistic control of land include, but are not limited to, the perpetuation of class and gender inequality, the removal of wealth from free market competition, the protection of individuals from a certain form of liability (individuals could not be alienated from an important asset), family feuding, and the concentration of political power within families across generations. For good reason, fee tails were eventually outlawed for real estate. In theory, what remained of attempts to control other forms of wealth from the grave were subject to the Rule against Perpetuities. Beginning in the sixteenth century, the state-enforced power of the dead hand extended only as far as a couple of generations with respect to real estate—and this was, at least in principle, true for all other assets.

Then, in the second half of the twentieth century, common law countries and jurisdictions began a slow process of repealing the Rule against Perpetuities.¹⁵ In the United States, the relevant date is 1995, when the state of Delaware repealed the Rule against Perpetuities for all assets other than real estate. Delaware was followed by Alaska in 1997, and this set off an arms race for states and financial institutions to repeal their own versions of the rule. The motivation for states was to attract high-net-worth individuals to the state's trust and banking systems. The primary motivation for wealthy individuals was that trust agencies had begun devising a new kind of trust to help people avoid taxes on wealth at death, and to avoid such taxes in perpetuity for all subsequent successors.¹⁶

14 Haskins, "Extending the Grasp of the Dead Hand."

15 Britain repealed it in 1964, Ireland in 2009, and South Australia in 1984.

16 Sitkoff and Schanzenbach, "Jurisdictional Competition of Trust Funds."

These repeals gave rise to the dynasty trust as we currently know it, which is in all respects the contemporary capitalist version of the fee tail.¹⁷

Dynasty trusts do not protect real estate, but they grant to the dead an equivalent—possibly a greater—power than the feudal fee tail. Assets that can be placed into dynasty trusts include things like stocks, bonds, shares in a hedge fund, and life insurance: assets that are easily convertible to cash, subject to compounding over generations, and, in the case of life insurance, can increase a trust's value tenfold upon the death of the testator.¹⁸ Not only are beneficiaries shielded from creditors as in a fee tail, they can also be protected by the dead hand from divorcee entitlements and obligations. Trust assets can be structured in such a way that they are in fact owned by the dead owner rather than subsequent generations, and so if a beneficiary crashes her car into a city building, the city cannot go after trust assets in a lawsuit, as they are not owned by the beneficiary.¹⁹ The testator can also place arbitrary conditions on whether and when beneficiaries can receive money. For instance, they can require that only heirs who enter into certain occupations will receive benefits, or that heirs who demonstrate unwanted behaviors like taking drugs, smoking, or converting to Catholicism, will not.²⁰ The dead can decide how much or how little power subsequent generations have to control parts of the trust.²¹ In effect, a dynasty trust is like a complicated conditional bequest—except that a dynasty trust allows the dead testator to earn unlimited amounts of money after death, and to conditionally bequeath such earnings for *each* subsequent generation, not just the one that follows after her own death.

In many ways, the dynasty trust has the potential to be at least as harmful as the medieval fee tail. The perpetuation of wealth and class inequality is a consequence of both. The dynasty trust, like the fee tail, creates a class of citizens who do not pay the same taxes as others do, who are not financially liable for wrongdoing as others are, and are therefore afforded state privileges and protections others are not. They are afforded such things at birth simply in virtue of the fact that a wealthy dead ancestor wished that this be so. On the other hand, beneficiaries are subject to coercive pressures if the trust requires antiquated or arcane conditions on obtaining benefits. True, the lucky descendants are free

17 Sitkoff and Schanzenbach, "Jurisdictional Competition of Trust Funds."

18 To see how this works, roughly, imagine that you purchase a life insurance policy that pays out \$10 million upon your death. Such a policy might be worth \$1 million now, and it can be placed in the trust to pay future beneficiaries.

19 Madoff, *Immortality and the Law*.

20 Ruud and Ruud, "Planning with the Dynasty Trust & Charity."

21 Ruud and Ruud, "Planning with the Dynasty Trust & Charity."

to refuse such benefits, but it is undeniably manipulative and maybe cruel to be confronted with the choice between a million dollars and the love of your life who happens to worship the wrong God. The dead are creating conditions that artificially incentivize their descendants to want to shape themselves according to the dead's wishes.

Finally, for those who care about capitalism and markets, dynasty trusts alienate large amounts of capital from the free market, subjecting it to the preferences and choices of someone who died long ago rather than those of individuals currently competing in the marketplace. It is no surprise that Adam Smith characterizes the dynasties of his day, entailed estates, as “barbarous” and “absurd” in *The Wealth of Nations*.²² A dynasty trust is a way to hoard capital, i.e., to take from the fruits of investment and hold it forever for disbursement to a small class of people at a comparatively miserly rate. Economic growth slows or shrinks when the living hoard rather than spend capital.²³ When the dead hoard money, we cannot justify it by saying that it is the free exercise of rational agents participating in the market. We must remember that the dead are not around to carry out their wishes. It is the state and court system that enforce the wishes of the dead. When the dead are allowed to hoard capital, what you have is the state intervening in the economy to actualize the preferences of dead people to hoard or spend, rather than the rational choices of living actors. There is no market-oriented justification for this practice.²⁴

We can generalize the argument by taking things to the limiting case. If the perpetual control of wealth is legitimate, then it is legitimate for an individual to perpetually refuse to alienate their wealth upon their own death. If earnings on assets rightfully belong to the dead, and must be spent or held according to their wishes, then given a sufficiently greedy set of high-worth people, there is a possible world in which it is legitimate for no property to be under the control

22 “But in the present state of Europe, when small as well as great estates derive their security from the laws of their country, nothing can be more completely absurd. They are founded upon the most absurd of all suppositions, the supposition that every successive generation of men have not an equal right to the earth, and to all that it possesses; but that the property of the present generation should be restrained and regulated according to the fancy of those who died perhaps five hundred years ago” (Smith, *The Wealth of Nations*, 384).

23 Keynes. *The General Theory of Employment, Interest and Money*; Rowe, “Keynesian Parables of Thrift and Hoarding.”

24 Haslett, “Is Inheritance Justified?” argues that asymmetric transfers like posthumous gifting violate equality of opportunity constraints of market capitalism, and are illegitimate as a result. Reich (*Just Giving*) argues that foundational, philanthropic giving, whether during life or death, is anti-market because such giving is unaccountable to consumer demand and shareholders, sells no goods, and has no marketplace competitors.

of anyone living. These *Maximal Scrooges* would be fully within their rights to ensure that the state enforce their wishes not to alienate any of their property to anyone. The only role for living people would be to protect the wealth of the dead from other living people. If this world is absurdly unjust, then it is unjust for the very reason that the principle of perpetual control is unjust.

I am not going to proceed any further in arguing that perpetual control is a bad thing. In this day and age, how much time do we need to make a case against feudal practices, whether on grounds of economic justice or libertarian free market principles? The fee tail was always unjust, and was never considered an instrument consistent with capitalism. But for some reason, the dynasty trust is. To the extent that it is, it is because of a principle of property rights that seems to be built into contemporary conceptions of property rights: freedom of testation.

1.2. Freedom of Testation

Philosophical justifications for free testation in the modern period date to Hugo Grotius, and can be found subsequently in the English tradition in John Locke and John Stuart Mill.²⁵ The justification takes for granted the modern conception of property rights. Whatever it means to legitimately acquire and own property, giving it away is one legitimate way to transfer such rights. Freedom of testation then likens testation to the power of gifting.²⁶ If I have a right to give away *X*, then I am free to give away *X* on some condition. One condition is a deferment. I can give my property away later, and after my death is certainly later. In addition, any potential recipient of a gift is free to refuse, and thus the offering of a gift, on whatever conditions, imposes no loss on a recipient. Since a gift can be deferred according to the wishes of the owner, a bequest after death is simply a deferred gift, and a conditional bequest simply an instance of a conditional gift. A dynasty trust or charitable trust, then, is simply a deferred conditional bequest with future generations or charitable purposes as the target recipient and the arbitrary stipulations in a deed are the conditions. Almost all of the practices I have out-

25 For an extensive discussion of the classical liberal responses to feudalism and inheritance of property rights, see Halliday, *Inheritance of Wealth*, ch. 2.

26 Grotius writes, "Aside from all positive law, intestate succession, as it is called, after ownership has been established, has its origin in natural inference as to the wishes of the deceased. Since the force of ownership was such that it could be transferred to another at the will of the owner, so also in case of retention of ownership at the time of death. . . . If any one had given no indication of his wishes, nevertheless, since it was not credible that his intention was to yield his property after his death to the first who would take it over, the inference is that his property is to belong to the person to whom it is especially probable that the dead man had wished that it should belong" (*The Rights of War and Peace, including the Law of Nature and of Nations*, 1625).

lined that honor the wishes of the dead in personal wealth can be derived from freedom of testation.

But the argument from freedom of gifting to freedom of testation presupposes that death does not alienate an owner from her rights to property. Get rid of this assumption, and freedom of testation cannot be derived. On the competing view that death alienates me from my property rights, I could no more give you my property after I die than I could give it to you after I have sold it to someone else, since selling is a way of alienating my rights. In life, we recognize selling, gifting, and some forms of confiscation (for civil liabilities, for instance) as legitimate alienations of property rights. We do not recognize death. But should we?

I do not think that it is an *intuitive* or *basic* assumption that death fails to alienate property rights. Most cultures and countries do not see freedom of testation as a natural or universal right. Primogeniture, for instance, has a much longer history as a practice in many cultures than free testation. In places and times where primogeniture has not held sway, inheritance practices seem to vary greatly depending on the social organization of a culture, whether it is misogynistic, agrarian, pastoral, colonial, and so forth.²⁷ In fact, Max Weber observed in the late nineteenth century that “complete, or nearly complete, liberty of testation is only recorded twice: as to Republican Rome and as to English Law; in both cases for expanding nations ruled by a landed gentry. Today the most important territory recognizing liberty of testation, is the territory of greatest economic opportunities: the United States.”²⁸ If Weber is right, or even close, then we have strong evidence that freedom of testation is a minority view, even in the West. Practices of primogeniture or inheritance, or communal “confiscation,” assume quite the opposite of the assumption that prevails in the United States today: they see some subset of the living as having rights to a person’s property following his or her death.

In fact, Locke’s ideas of individual property rights, from which the English and American conception of property derives, are famously responses to Filmer and Filmer’s extended justification of feudal monarchic practices where individuals could no more stipulate rights of inheritance than kings could stipulate successors to the crown.²⁹ Filmer famously derived *inheritance rights*, not freedom of testation, from the Abrahamic creation myth of Adam and his sons. This is to say that even among Englishmen of the recent past, free testation is far from intuitive. It would be surprising that an intuitively obvious assumption happens to be rejected by close to 100 percent of human cultures and history.

27 Murdock, “Ethnographic Atlas Codebook”; Nussbaum, “Liberty of Testation.”

28 Weber, *The Theory of Social and Economic Organization*.

29 Filmer, *Patriarcha*.

What can justify the assumption that death does not alienate me from my property rights, if not intuition? Even in Locke, who advocates freedom of testation, we cannot find a justification. Locke famously argued that nature was the common stock of mankind bestowed on humans by God, and that property rights arose from individuals mixing their labor with nature. Locke foresaw that one possible consequence of this view is that, at death, property returned to “the common stock of mankind.” In response, Locke mentions freedom of testation as a natural right associated with the right of property—but never defends it. Instead, he makes an extensive case for a child’s right to inheritance on the grounds that parents have an obligation to care for their children. As we have seen, though, rights to inheritance are incompatible with free testation.³⁰ If you are entirely free to dispose of your property at your will, then your children do not have a right of inheritance unless this is granted by your testament. Locke’s stated view that there is a natural right to free testation is hard to square with his argument for children’s right to inheritance.³¹ In defending a right to inheritance, Locke in fact opens the door for arguments that moral obligations derived from considerations wholly outside of the wishes of the dead determine the just distribution of wealth. These lines of reasoning have a subsequent history in Western philosophy: Hegel, for instance, explicitly makes the case against freedom of testation on broadly Lockean grounds, arguing that obligations to family alone determine the parameters of morally defensible distribution of wealth after death.³²

Another source for the modern concept of freedom of testation comes from John Stuart Mill. In his *Principles of Political Economy*, Mill claimed that free testation, as opposed to inheritance, is analytic to the very concept of private property.³³ But even Mill eventually recognized that this is not a tenable claim.³⁴

30 Waldron, “Locke’s Account of Inheritance and Bequest.”

31 Waldron, “Locke’s Account of Inheritance and Bequest.”

32 Hegel, *The Philosophy of Right*, 216–18.

33 Mill, *Principles of Political Economy*, bk. 2, ch. 2, sec. 3: “Nothing is implied in property but the right of each to his (or her) own faculties, to what he can produce by them, and to whatever he can get for them in a fair market; together with his right to give this to any other person if he chooses, and the right of that other to receive and enjoy it. It follows, therefore, that although the right of bequest, or gift after death, forms part of the idea of private property, the right of inheritance, as distinguished from bequest, does not. That the property of persons who have made no disposition of it during their lifetime, should pass first to their children, and failing them, to the nearest relations, may be a proper arrangement or not, but is no consequence of the principle of private property.”

34 Mill revised his view so that, by the posthumous publication of *Chapters on Socialism*, he takes almost nothing to be analytic to the concept of private property. “The idea of property is not some one thing, identical throughout history and incapable of alteration, but is variable like all other creations of the human mind; at any given time it is a brief expression

A community that employed all of our current forms of property protection, including freedom of gifting, but decided that, upon death, property would be inherited by family or divided equally among all community members (or relinquished to the state), would not thereby fail to share our concept of private property. The members of this imagined community simply have a disagreement with us as to how far it extends. One can consistently believe in private property and believe that, upon death, moral considerations outside of one's own wishes determine the distribution of one's assets. That individuals have freedom of testamentation is a substantive assumption, not an analytic truth.³⁵

In the end, I believe two considerations speak against the idea that death preserves property rights. The first is the consideration that with death comes the end of the subject, and therefore an end to the person with claims of rights and property. Upon death, there is no subject whose rights we are violating when we take a piece of property that was once theirs and do with it something contrary to their wishes. This is the consideration I address extensively in the final section, where I argue against the existence of posthumous harms. The other consideration is that I do not believe it is possible to consistently maintain that one's property rights survive death, but one's political rights do not. Even more importantly, it is not possible to maintain that one's property rights with respect to some kinds of property, like personal assets, survive death, but one's property rights with respect to another kind of property, like business assets, do not survive death.

I have already given reasons, in the first section, for claiming that a right of suffrage does not entail a right of posthumous suffrage, nor does freedom to make decisions with respect to a business I own entail the right to such decisions after my death. The question is whether there is a relevant difference between these political and economic freedoms, afforded to us only while we are alive, and our freedom of gifting associated with private property. I believe that there is no relevant difference.

Political rights, like property rights, can be given away conditionally. I can agree in a heavily Republican state to vote for Ralph Nader on the condition that a Nader supporter in a swing state vote for Al Gore. Two senators can agree to support each other's bills, which they otherwise oppose, as part of a deal. Such vote pairing, or agreements to exchange votes strategically, are quite a common

denoting the rights over things conferred by the law or custom of some given society at that time; but neither on this point nor on any other has the law and custom of a given time and place a claim to be stereotyped for ever," *Fortnightly Review*, 135–36.

35 For an extensive defense of Mill's considered views on inheritance and taxation, see Cappelletten and Pederson, "Just Wealth Transfer Taxation."

occurrence in parliaments, and not at all uncommon in general elections. Similarly, business rights can be given away conditionally. I am perfectly within my rights to conduct my business strategically according to your wishes, or offload certain business decisions on other managers at a firm. Political and private business decisions are freely and conditionally given.

Rights that I hold in political and business contexts can also be similarly made into deferred gifts, as when I promise to vote for the next three Democrats on the condition that you vote for the Republican in the next election, or when I agree that I will manufacture my product in your town in the next ten years on the condition that you lower sanitation taxes. Of course there are legal and moral limits to gifting, but this is true with respect to our private property as well. I cannot give away my property to enemies of the state, for instance, or hire people to perform illegal activities in my life or my business. The state is similarly required to enforce my liberties across all of these activities during my lifetime; a state infrastructure prevents others from suppressing my vote and political activity, and civil courts enforce the contracts I make for my private business. Similarly, civil courts and law enforcement are meant to protect my private property rights.

Finally, I have stakes in the future of political institutions and my business as much as I have a stake in the future of my personal wealth. I help to erect and change political institutions with my votes and political activity. The future of a business I found and work hard to build can be even more important to me than what happens to my house. So in almost all respects relevant to their status as freely and conditionally transferable powers, my political rights, private business rights, and private property rights are the same. And yet we treat two of these as alienated at death, for good reasons, but one of them as preserved, for no discernible reason.

1.3. An Objection and Reply

One difference between our rights to private assets and our business assets is worth discussing. This difference threatens the argument I have been making. In some jurisdictions, you are not allowed to discriminate against certain people with respect to doing business with them as a business owner, but you are permitted to discriminate against them as a customer, i.e., spender of private wealth. For example, you cannot as a baker refuse to bake a cake for a gay wedding, but you are perfectly within your rights to refuse to purchase a cake from a business that supports gay marriage. One moral principle underlying this legal practice may be that we view business rights as more restricted than private property rights, and that this is the explanation as to why private property rights survive death but business rights do not. The view might be that private businesses inter-

act more directly with public needs and public policies than individual private property, and that private businesses benefit more from public goods than private property, therefore giving the government justification for imposing limits on freedom in one case but not the other. Thus, the reason for prohibiting posthumous rights to business activity is because of the need to protect public goods that businesses provide but private wealth does not. On this view, we restrict posthumous business rights for public policy reasons that do not carry over to private wealth.

The reasoning here is unpersuasive to me. First, it is highly contentious whether there is a moral, rather than a practical, justification for treating businesses differently from individuals with respect to rights to discriminate. Organized consumer boycotts that wrongfully discriminate could just as well have public policy implications at least as great as those threatened by businesses that wrongfully discriminate, so the justification on the grounds of policy implications fails. The truth is, the state is actually able to enforce unjust discrimination among businesses, at least some of the time. There is no (wieldy) way to enforce claims against customers for unjust discrimination.

Second, the appeal to public goods as the justification for differential treatment is unpersuasive. Higher-net-worth individuals—those who exercise the most power due to their private property rights before and after death, benefit greatly from public, tax-funded goods, such as financial laws and federally backed insurance, the enforcement of financial regulation, the Federal Reserve, government-backed Treasury bonds, the Securities and Exchange Commission, and the highly expensive litigation system that enforces contracts at all levels of government. In the end, if one believes it is just for states to enforce what amounts to a moral prohibition on businesses that wrongly discriminate, it stands to reason that it is just for states to enforce a prohibition on customers who wrongly discriminate. Here, I am in agreement with Libertarians, who accept the contrapositive.

Turning to posthumous rights, I similarly believe there is no way to justify the denial of posthumous rights to political and business decisions on the grounds of public interest without denying such rights to private wealth on the same grounds. Here I am assuming no more of “public interest” than that it means “in the interest of the public.” There is no sense of “public interest” in which bequests of private wealth are not privy to considerations of public interests, but posthumous wishes to vote or conduct business are. When large amounts of personal wealth can be squandered, hoarded, or given to charities for guinea pigs, this is as much in the public interest as when some dead industrialist wanted to keep manufacturing in a certain town. Similarly, the distribution of funding

in the charity sector is certainly a matter of public interest. The charity sector in many ways is designed to supplement and complement failures of the state and private sector to provide for certain goods. Who has power and how much power they have in these sectors, before and after death, are as much a matter of public interest as they are in these other sectors.

The treatment of private wealth on the one hand, and business wealth and political capital on the other, is highly artificial. There is no sense in which the machine I purchase for my business is subject to one set of property rights that expire at death, and the dollar that I earn from that machine is subject to another set that is subject to freedom of testation. If the very idea of property rights requires a right to free testation, i.e., such rights are not alienated with the death of the subject, then it should require posthumous rights to business activity. If public interest considerations overrule any felt need to honor the wishes of the dead in businesses and politics, it should overrule the same need for private wealth as well.

2. POSTHUMOUS HARM

The philosophical justification for honoring the wishes of the dead arises from considerations of *posthumous harm*, something that Adam Smith recognized quite early in his writings criticizing perpetual entailed estates.³⁶ The argument for the existence of posthumous harm rests largely on a few kinds of common-sense cases. For instance, it seems wrong to spread malicious lies about someone who is now dead.³⁷ Second, it seems that if an author greatly wanted a book of theirs to be published, it would be wrong to throw away the manuscript now that she is dead. Finally, it seems that if you made a promise to give a dying man's locket to his grandson, it is wrong not to do so.³⁸ These things seem wrong, even though there is no longer a subject of harm, like there is no subject who has property rights to be violated. These kinds of cases suggest that, even though the subject of harm is now dead, we still feel that some kind of harm or wrong occurs. The central philosophical problem is characterizing who has been harmed, or whose rights are violated, if the wishes of the dead are thwarted. The philosophical moves begin: perhaps not all wrongs require a subject; perhaps the sub-

36 Smith writes: "The difficulty is to find at what period we are to put an end to the power we have granted a dying person of disposing of his goods. . . . The insensible progress of entails was owing to their not knowing how far the right of the dead might extend, if they had any at all" (*Lectures on Jurisprudence*, 165–69).

37 Feinberg, *The Moral Limits of the Criminal Law*.

38 Wisniewski, "What We Owe the Dead."

ject is the premortem person; perhaps there is a sophisticated form of backward causation; perhaps the best language is not in terms of “harms” or “rights,” but rather a matter of welfare and well-being, where a person’s well-being during life depends on factors in the future.³⁹ Either way, the intuitions must be preserved, even though the right philosophical theory is elusive.

To me, the commonsense cases are not particularly persuasive. Certainly, they are not persuasive enough to be foundational in a theory that grants the dead posthumous rights. First, most of the intuitive cases involve one-off, particular, and context-specific moral obligations that individual persons have to other recently dead persons. The generalizability of these kinds of cases to cases involving perpetual, *institutional* obligations to the dead is questionable.⁴⁰ Second, consider the case of maliciously lying about the dead. Is it really wrong to spread a lie about a dead person if you stipulate that no living member of her friends or family has any stake in the dead person’s reputation, and you also control for any judgment on the categorical wrongness of lying or for the objectionable states of mind or character manifest in lying about the dead? I think that if you do control for these things—ensuring that no one living cares about the lie, that the lie is not an expression of an objectionable state of mind, and that it represents a case of permissible lying—there is no wrongness left to explain. Deathbed promises are the final type of case. These promises are neither unconditional nor eternal, nor must they be satisfied at serious financial or moral cost to the living. They are, instead, a lot like living promises, which can be overridden by the circumstances of morality.⁴¹ If I promise my child some candy, but through no fault of my own the only available candy must be acquired at serious moral cost to some current candy owner, it is not morally obligatory to fulfill this promise. A promise never holds overriding moral weight. And again: Is thwarting a promise to the dead wrong in virtue of having thwarted the wishes of the dead, or because of some kind of categorical judgment on the state of mind or character of a promise breaker? Control for these variables, and I do not think it is clear that there is some intuition of harm to the dead.

Despite this skepticism, I do accept that there is something bad about thwarting someone’s future-facing desires, projects, and values. This I concede to the advocates of posthumous harms. But I believe the badness does not arise from harm. Many of our most significant values depend for their realization on the ex-

39 See Feinberg, *The Moral Limits of the Criminal Law*; Pitcher, “The Misfortunes of the Dead”; Levenbook, “Harming Someone after His Death”; Wisniewski, “What We Owe the Dead”; and Portmore, “Desire Fulfillment and Posthumous Harm.”

40 Thanks to an anonymous reviewer for this consideration.

41 This point is made in Wisniewski, “What We Owe the Dead.”

istence of future people, and the ways our own activities in life affect them. This is a central principle in Scheffler's *Death and the Afterlife*. Valuing for Scheffler is constitutively connected with conserving what is valued in a future world inhabited by human beings where such values are manifest. If the human race were to go extinct shortly after our own deaths, many of our projects would turn out not to matter anymore.⁴² The young adult novel I wrote, the cure for cancer I helped advance, the electoral college reform I fought very hard to push: these will all turn out not to matter if, shortly after my death, humans go extinct. It is a very short step to the idea that, even if humans continued to persist but somehow my life projects were simply erased at my death, I would similarly have engaged in insignificant or pointless activities, as my values would no longer be preserved and perpetuated. One of the most important values that gets undermined in destroying my life's work upon my death is my mark on the future, something that Scheffler convincingly argues is more significant to me, as someone with values, than even my own life. The amount of meaning, value, and significance of much of my life's work depends on its impact on the future of humankind. If there is no humankind, or if the future consists of a human race where I make no mark, or have no impact, there is a sense in which my life and life's activities have been made pointless.

Doug Portmore has called the desire for current projects and values that are future directed to be preserved after death "not-for-naught" desires.⁴³ You would like your activities to be meaningful and impactful. You do not want them to be *for naught*. When those activities depend for their fulfillment on events in the future, a future under the control of future generations, then future people determine whether your work will be for naught. Because I do not want my activities to be for naught, it makes sense that I aim to set up institutions that help to guarantee a place for them. Hence, we have institutions that enforce the wishes of the dead.

Not-for-naught reasoning is pervasive even during our lives. Thomas Kelly shows that many forms of purportedly fallacious "sunk-cost" reasoning are in fact a legitimate form of not-for-naught reasoning.⁴⁴ Geffen Records continued to fund the Guns N' Roses album *Chinese Democracy* for almost two decades,

42 Scheffler actually defends three different theses here: that things *mattering to us* depends on *our confidence* in the existence of a future human race, that things *mattering* depends on the existence of a future human race, and that things *mattering to us* depends on *the existence* of a future human race.

43 Portmore, "Desire Fulfillment and Posthumous Harm," and "Welfare and Posthumous Harm."

44 Kelly, "Sunk Costs, Rationality, and Acting for the Sake of the Past."

putting millions upon millions of dollars into the budget of a failing album in hopes that all of the time and money that went into the production would eventually materialize into a bestseller. In the justification for continued US involvement in failing counterinsurgency campaigns in Vietnam, then Iraq and Afghanistan, there has often been an appeal to the idea that troops would have “died for nothing” if the United States abandoned the campaign at a given moment, with the foreseeable side effect that the host nation would fall into enemy hands shortly after a US withdrawal. Kelly persuasively argues that not-for-naught reasoning can very well be rational if projects turn out as intended in the long run.⁴⁵ The future can just as well render projects meaningful and worthwhile as it can render them for naught.

There is a sense in which thwarting an individual’s wishes after their death with respect to their conditional, charitable, or dynasty bequests undermines the perpetuation of their values into the future, values that are forward looking and depend on their realization on the fact that there is an intended impact on future generations. Milton Hershey of the Hershey Company put all of his money and his entire corporate empire under a trust to fund an orphan school of one hundred poor white male orphans, in perpetuity.⁴⁶ If you decide that, upon Milton Hershey’s death, his life mission of funding an orphan school is no longer enforceable by the state, and the board of his company decides instead to invest in organic milk production, you are undermining one of Hershey’s forward-looking interests and values, just like you would be undermining an interest if you just burned a book manuscript a dying person wishes to be published. But does this make the thwarting, or rendering of a particular project of a dead person “for naught,” a posthumous harm, or a wrong to that individual such that they have a right that we not do this?

I believe the answer is no. The mere making of a project “for naught” is not sufficient to require state enforcement of a posthumous wish. We do not have a right that individuals in the future not render our lives less significant than we would like. If our wishes and activities turn out to be in the better interests of future lives, then those wishes may continue to be honored simply because the living can recognize this and do what is best for themselves, and thus my life remains significant. But my life will be significant for the right reason, namely, because my wishes and activities are in fact good for the future world. The wishes of dead me should be honored precisely because they are the right ones to

45 Lest you mistake my meaning, I am certainly not saying that *Chinese Democracy* or the wars in Vietnam and Iraq, three of the greatest disasters of recent American history, are examples of such rational projects.

46 This is the primary case study in Lam, “The Wishes of the Dead.”

have for the living, and not because they are mine. Future people are not obliged to see to it that my wishes are honored so that the significance of my life may be maximized.

But I do not know whether these assertions amount to an argument. Some have a strong intuition that the rendering of people's projects insignificant is wrong. Many even consider it a backward-looking harm, affecting the pre-mortem person. Portmore claims as much. If it is wrong, I believe it is an "impersonal wrong." There is something bad in the world with having people lead insignificant lives, but it is no harm to the person who leads this life. After all, probably lots of people have, and continue to lead, rather insignificant lives, and do not feel the worse for it. None of their values are likely to be perpetuated even in the immediate future after their death. They will make zero impact. I do not think it is harmful to these people that we do not actively make their lives more meaningful by seeking out and realizing their values.

But I am willing to concede the point for argument's sake. Perhaps it is better, all things being equal, to have in the world people who lead more significant rather than less significant lives. Perhaps it is some form of harm to the pre-mortem person that we do things that render their lives less significant. Even with these concessions, I will now argue, the harm of being rendered less significant is not sufficient to justify the state enforcement of posthumous wishes.

Consider the following scenario: a certain author has spent his entire life finding the homeopathic cure for pneumonia, an affliction that claimed his dear mother. His book has been finished and it is his dying wish that it be published. Upon reading the book, living individuals are of the judgment that it is misguided to the point of being dangerous, and that it serves no interests of the living to have this book published. But out of respect for the dead man, they publish it anyway. Years down the line, people seeking homeopathic cures buy and use the book, and hundreds die of pneumonia as a result of not seeking standard treatments. (But as it is with all things homeopathic, people tend never to blame the book, and its influence persists.) It turns out that the result of the publication served no interest of any subsequent living individual, and in fact caused a great deal of harm.

I do not think this particular case was decided morally or prudently. Thwarting a person's valued projects can make her life less significant, to be sure—but sometimes implementing them can make her life (at least as reflected by these particular projects) positively unjust. If the former is a wrong future people can inflict on past people, then so too is the latter. The relevant principle I believe is true is that, if it is wrongful or harmful for future people to render a dead person's projects "for naught," then it is at least as wrongful or harmful for future people

to render a dead person's projects "for wrong" or "for worse." If the future has the power to render a past project significant or insignificant, and this is a right or wrong to the individual, then the future also has the power to render such a project wrongful, and this too is a wrong to the individual.

Ultimately, it is the living who know about the effects of thwarting or implementing the wishes of the dead, since they will be the ones implementing or thwarting such wishes in a world that is quite different from the one the dead person occupied. So it will be the living who make judgments of harm and justice. If this is true, the decisions of the living to thwart or implement the wishes of the dead reduce to the decision of what policy is the best thing for the world they inhabit. If a policy is good to do, then they should do it because it is good for the living; it is fortuitous that some dead person also wanted their wealth put to this purpose. If a policy is bad to do, then they should not do it, for the fact that some dead person wanted it does not change the fact that it is bad, and doing bad in the name of a dead person is wrong or harmful to the dead person, under the hypothesis that there is posthumous harm. The fact that a wish is a wish of the dead is no moral reason to implement that wish.

As a matter of fact, there are "public policy" constraints on conditional and charitable bequests in the law just as there are "public policy" constraints on gifting. You cannot freely testate that the works of Rembrandt in your collection be set on fire, or that your fortune be donated to a drug cartel. But current public policy constraints only go so far; they do not protect the living from bequests that require them to marry or raise children within a certain religious faith, for instance, and it is still possible to word a bequest carefully enough so as to incentivize an inheritor to divorce a spouse the testator did not like.⁴⁷ There is far too much control preserved for the dead even with acknowledgment that there are some public policy constraints on testation.

In the law concerning charitable giving, a topic that I have not discussed at length here but have discussed extensively in season 1, episode 1 of the *Hi-Phi Nation* podcast, there is a doctrine called *cy-pres* that states that, should a charitable purpose now violate the law or public policy, organizations are still legally obligated to spend money as close as possible to the wishes of the original testator consistent with public policy.⁴⁸ This means, for instance, that the Hershey school can no longer discriminate against nonwhite children, since that is now illegal. But public policy does not include considerations of overall social fairness and distribution of resources. If an enormous amount of money donated to a university has as a condition that the money be spent on parapsycholog-

47 *In Re Estate of Feinberg*; Madoff, *Immortality and the Law*.

48 Lam, "The Wishes of the Dead."

ical research, but a small amount of money to the same organization has as a condition that it is spent on antibiotic-resistance research, it is of great moral significance that such money is distributed absurdly out of line with the pressing needs of today. But it is of no legal significance in terms of “public policy” provisions. So public policy is important. In fact, it is of ultimate importance, and this is certainly not reflected in the law, and we have bad philosophical ideas about posthumous rights to thank for this.

3. CONCLUDING THOUGHTS

In sum, I do not believe that the dead can be wronged or harmed in virtue of our rendering of their projects less significant. Thus, they do not have a right that we honor their posthumous wishes with respect to their personal wealth. However, even if we accept that they can be so harmed, those considerations lead us to conclude that they can equally be wronged or harmed if their projects are allowed to be carried out, by the living, for harmful or unjust ends, ends that the dead are unable to see and change today because they are dead. As a result, even considerations of posthumous harm do not justify the practice of honoring the wishes of the dead contrary to the needs and values of the living.

Before I conclude, I want to mention one more kind of justification I have not discussed in this paper, which concerns consequentialist justifications for freedom of testation, perpetual control, and the granting of posthumous legal rights. These views state that the existence of these things incentivizes people in desirable ways. Perhaps we need freedom of testation so that people can compete in the market for a wealthy person’s bequests. Perhaps perpetual control incentivizes forward-looking, non-egoistic, and selfless concerns about the future. The idea is that if we do not allow people to become stakeholders in the future of our world, we will see a bunch of selfish egoists wantonly destroying the planet.

I have not considered these justifications because I believe that the denial of free testation, perpetual control, and posthumous rights is on quite solid footing in terms of consequentialist reasoning. The idea that we need free testation so that children will treat their parents better in an effort to win a bequest is a perversity only an economist could conjure up. To say that a world of aristocratic dynasties is better than a world in which people begin with the same set of rights and responsibilities is about as persuasive as saying that Britain during feudalism and fee tails was better than Britain today. In the charity sector, perpetual foundations simply do worse empirically than sunset foundations in producing good outcomes.⁴⁹ The idea that fixed-purpose charities that last forever devised by a

49 Madoff, *Immortality and the Law*, ch. 3.

person who is long dead will do more net good than charities that respond to the concerns of the best interests of the living is analytically false in any world in which charitable needs change over time.

Ultimately, we the living are to blame for sabotaging our own best interests. The dead are not around to complain if we were to stop honoring their wishes. These are our institutions, and any pain we inflict on ourselves from being worse off but for the preferences of the dead cannot honestly be held against the dead. The best we can say on behalf of our practices is that we have a self-interested desire that our interests and values not completely vanish from the world after we die, and so we erect institutions against our current interests so as to protect our future interests from future people who may undermine them. Honoring the wishes of the dead, then, is a mere side effect of our own vanity and narcissism in seeking honor when we are dead. It is time we break this cycle of absurdity, as we have in the past, and leave the best legacy we can for the living, which is the autonomy to determine the best world to make for themselves.

Vassar College
balam@vassar.edu

REFERENCES

- Baker, John. *The Oxford History of the Laws of England*. Vol. 7, 1483–1558. Oxford: Oxford University Press, 2003.
- Cappelen, Cornelius, and Jørgen Pedersen. “Just Wealth Transfer Taxation: Defending John Stuart Mill’s Scheme.” *Politics, Philosophy and Economics* 17, no. 3 (2018): 317–35.
- Feinberg, Joel. *The Moral Limits of the Criminal Law*. Vol. 1, *Harm to Others*. Oxford: Oxford University Press, 1987.
- Filmer, Robert. *Patriarcha, or the Natural Power of Kings*. London: Richard Chiswell, 1680.
- Grotius, Hugo. *The Rights of War and Peace*. Translated by A. C. Campbell. New York: M. Walter Dunne, 1901.
- Halliday, Daniel. *Inheritance of Wealth: Justice, Equality, and the Right to Bequeath*. Oxford: Oxford University Press, 2018.
- Haskins, George L. “Extending the Grasp of the Dead Hand: Reflections on the Origins of the Rule Against Perpetuities.” *University of Pennsylvania Law Review* 126, no. 19 (1977): 19–46.

- Haslett, D. W. "Is Inheritance Justified?" *Philosophy and Public Affairs* 15, no. 2 (Spring 1986): 122–55.
- Hegel, G. W. F. *The Philosophy of Right*. Edited by Allan W. Wood. Cambridge: Cambridge University Press, 1991.
- Jefferson, Thomas. Thomas Jefferson to Major John Cartwright, June 5, 1824. In *The Writings of Thomas Jefferson*, vol. 16, 42–52. Washington, DC: Thomas Jefferson Memorial Association, 1903.
- Kelly, Thomas. "Sunk Costs, Rationality, and Acting for the Sake of the Past." *Nous* 38, no. 1 (March 2004): 60–85.
- Keynes, John Maynard. *The General Theory of Employment, Interest and Money*. London: Macmillan, 1936.
- Lam, Barry. "The Wishes of the Dead." January 31, 2017. In *Hi-Phi Nation*, season 1, episode 1. Podcast, 49:28. <https://hiphination.org/complete-season-one-episodes/episode-one-the-wishes-of-the-dead>.
- Levenbook, Barbara Baum. "Harming Someone after His Death." *Ethics* 94, no. 3 (April 1984): 407–19.
- Luper, Steven. "Posthumous Harm." *American Philosophical Quarterly* 41, no. 1 (January 2004): 63–72.
- Madoff, Ray D. *Immortality and the Law: The Rising Power of the American Dead*. New Haven: Yale University Press, 2010.
- Mill, John Stuart. *Chapters on Socialism*. Toronto: University of Toronto Press, 1879.
- . *Principles of Political Economy*. London: John W. Parker, 1848.
- Murdock, George P. "Ethnographic Atlas Codebook." *World Cultures* 10, no. 1 (1998): 86–136.
- Nussbaum, Arthur. "Liberty of Testation." *American Bar Association Journal* 23, no. 3 (March 1937): 183–86.
- Otsuka, Michael. *Libertarianism without Inequality*. Oxford: Oxford University Press, 2003.
- Paine, Thomas. *Rights of Man: Being an Answer to Mr. Burke's Attack on the French Revolution*. London: J. S. Jordan, 1791.
- Picketty, Thomas. *Capital in the Twenty-First Century*. Cambridge, MA: Belknap Press, 2014.
- Pitcher, George. "The Misfortunes of the Dead." *American Philosophical Quarterly* 21, no. 2 (April 1984): 183–88.
- Portmore, Douglas W. "Desire Fulfillment and Posthumous Harm." *American Philosophical Quarterly* 44, no. 1 (January 2007): 27–38.
- . "Welfare and Posthumous Harm. Unpublished manuscript. http://www.public.asu.edu/~dportmor/Welfare_and_Posthumous_Harm.pdf.

- Reich, Rob. *Just Giving: Why Philanthropy Is Failing Democracy and How It Can Do Better*. Princeton: Princeton University Press, 2018.
- Rowe, Nicholas. "Keynesian Parables of Thrift and Hoarding." *Review of Keynesian Economics* 4, no. 1 (January 2016): 50–55.
- Ruud, Judith K., and William N. Ruud. "Planning with the Dynasty Trust and Charity." *Gift Planner's Digest*, January 4, 2000. <https://www.pgdc.com/pgdc/planning-dynasty-trust-charity>.
- San Antonio Express-News*. "Widow Fumes at Order to Start Smoking." September 10, 1993.
- Scheffler, Samuel. *Death and the Afterlife*. Oxford: Oxford University Press, 2013.
- Sitkoff, Robert H., and Steven R. Horowitz. "Unconstitutional Perpetual Trusts." *Vanderbilt Law Review* 67, no. 6 (2014): 1769–1822.
- Sitkoff, Robert H., and Max M. Schanzenbach. "Jurisdictional Competition of Trust Funds." *Yale Law Journal*, 115, no. 2 (2005): 356–437.
- Smith, Adam. *Lectures on Jurisprudence*. Edited by R. L. Meek, D. D. Raphael, and P. G. Stein. Oxford: Oxford University Press, 1978.
- . *The Wealth of Nations*. Edited by Edwin Cannan. New York: Modern Library, 2000.
- Waldron, Jeremy. "Locke's Account of Inheritance and Bequest." *Journal of the History of Philosophy* 19, no. 1 (January 1981): 39–51.
- Weber, Max. *The Theory of Social and Economic Organization*. Edited by Talcott Parsons. New York: Free Press, 1947.
- Wisniewski, J. Jeremy. "What We Owe the Dead." *Journal of Applied Philosophy* 26, no. 1 (February 2009): 54–70.

REVISITING THE ARGUMENT FROM ACTION GUIDANCE

Philip Fox

THE PHILOSOPHICAL LITERATURE witnesses a by now fairly large debate about the degree to which normative notions like “ought” or “reason” are perspective dependent. Roughly speaking, *objectivists* believe that what one ought to do depends on all the facts, whereas *perspectivists* believe that what one ought to do depends instead on one’s epistemic perspective on the facts (and so only on facts that are epistemically available in some yet-to-be-specified sense).¹ Their dispute traditionally revolves around examples like the following:

Doctor: Jill, a doctor, must decide how to treat her patient. She only has two options: prescribe pill A or pill B. After careful research, her evidence decisively suggests that A is the cure and B is lethal. In fact, however, the reverse is true: A is lethal and B is the cure. Jill knows that the patient will die soon if she does nothing and that she has no time for further research. What ought Jill to do?²

- 1 For proponents of perspectivism, see, e.g., Andrić, “The Case of the Miners,” “Objective Consequentialism and the Licensing Dilemma,” and “Objective Consequentialism and the Rationales of ‘Ought’ Implies ‘Can’”; Jackson, “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection”; Kieseewetter, “‘Ought’ and the Perspective of the Agent,” “How Reasons Are Sensitive to Available Evidence,” and *The Normativity of Rationality*; Lord, “Acting for the Right Reasons, Abilities, and Obligation”; Prichard, “Duty and Ignorance of Fact”; Robertson, “Epistemic Constraints on Practical Normativity”; and Zimmermann, *Living with Uncertainty and Ignorance and Moral Obligation*. For proponents of objectivism, see, e.g., Bykvist, “How to Do Wrong Knowingly and Get Away with It”; Graham, “In Defense of Objectivism about Moral Obligation”; Littlejohn, “Do Reasons and Evidence Share the Same Residence?”; Moore, *Ethics*; and Thomson, “Imposing Risks” and *Normativity*.
- 2 This example is taken from Kieseewetter (“‘Ought’ and the Perspective of the Agent”), who presents it as a simplified version of a case by Jackson (“Decision-Theoretic Consequentialism and the Nearest and Dearest Objection”). Similar examples are discussed in Andrić, “Objective Consequentialism and the Licensing Dilemma”; Kieseewetter, *The Normativity of Rationality*; Graham, “In Defense of Objectivism about Moral Obligation”; McHugh and

The answer is not obvious. On the one hand, one could side with the objectivist and say that Jill ought to prescribe *B*. After all, this is the only way of saving the patient's life. On the other hand, one could side with the perspectivist and say that Jill ought to prescribe *A*. After all, choosing any other option seems completely irresponsible.

What makes cases like *Doctor* interesting is that both positions seem to have significant intuitive appeal. Even if one is initially more attracted to one of them, it is easy to imagine how someone might have ended up on the other side. Partly for that reason, there is little ground for thinking that intuition alone will help advance the debate between objectivists and perspectivists much further. At best, appealing only to our intuitions looks like a recipe for a stalemate.

Fortunately, a less intuition-based, more theory-driven approach is available. In order to pursue such an approach, this paper draws together two separate themes from the recent literature: first, that the normative is in some sense action guiding, and second, that a discussion of perspective dependence in the practical domain benefits from a discussion of similar issues in the epistemic domain.³ The main thesis is that objectivism is incompatible with a very plausible assumption about the possibility of being correctly guided by a normative theory.

Put succinctly, my argument against objectivism is as follows:

The Argument from Action Guidance

1. It is sometimes possible for a normative theory to correctly guide action.
2. If objectivism is true, this is never possible.
3. Therefore, objectivism is false.

The central advantage of this argument is that it only makes an extremely weak assumption about the action-guiding role of normative theories. In particular, it does not assume that normative theories are *always* capable of guiding action, and—as I will explain in more detail below—it thereby avoids recent criticisms of guidance-based arguments against objectivism.

With the main argument on the scene, let us pause briefly for a general obser-

Way, "Objectivism and Perspectivism about the Epistemic Ought"; and Way and Whiting, "Perspectivism and the Argument from Guidance."

3 For the first theme, see, e.g., Graham, "In Defense of Objectivism about Moral Obligation"; Kiesewetter, "How Reasons Are Sensitive to Available Evidence"; Lord, "Acting for the Right Reasons, Abilities, and Obligation"; Sepielli, "Subjective Normativity and Action-Guidance"; and Way and Whiting, "Perspectivism and the Argument from Guidance." For the second theme, which has received less attention, see, e.g., McHugh and Way, "Objectivism and Perspectivism about the Epistemic Ought."

vation about arguments of this kind. I take it that all guidance-based arguments face the same choice. First, they could employ a fairly demanding notion of action guidance, in which case it will be relatively straightforward to explain why objectivism is inconsistent with it, but also correspondingly more difficult to show that this notion does not beg the question against objectivists right away; this approach is taken by Errol Lord, and it has been criticized accordingly by Clayton Littlejohn and Jonathan Way and Daniel Whiting for its strong assumptions about the possibility of action guidance.⁴ Or second, guidance arguments could employ a fairly modest notion of action guidance, in which case it will be relatively straightforward to show why it does not beg the question against objectivists, but correspondingly more difficult to explain why objectivism is inconsistent with it.⁵ This paper takes the second approach, and so for the most part I will be concerned with defending premise 2 above.

I proceed as follows. Section 1 introduces objectivism and perspectivism as two substantive views about the perspective dependence of “ought.” In section 2, after setting out two central assumptions of this paper, I develop and defend an account of what it is for a normative theory to correctly guide action. Section 3 anticipates a number of potential objections to this account. In section 4, I use this account to explain why objectivism rules out that normative theories are ever correct guides to action. Section 5 concludes. Throughout, I aim to advance the debate over the perspective dependence of “ought” beyond merely intuitive considerations and to shed light on a notion of action guidance that should be of independent interest to normative theory.

1. OBJECTIVISM AND PERSPECTIVISM

I have already mentioned *Doctor*, a representative example that illustrates the disagreement between objectivists and perspectivists. To repeat, Jill has two pills available to treat her patient: pill A, which would be lethal, and pill B, which would cure the patient. She also has decisive but misleading evidence about which is which. What ought Jill to do?

Objectivists will say that Jill ought to prescribe B, because, on their view, what one ought to do depends on all the facts. This includes the fact that pre-

- 4 Lord, “Acting for the Right Reasons, Abilities, and Obligation”; Littlejohn, “Prichard’s Question and Reasons Perspectivism”; Way and Whiting, “Perspectivism and the Argument from Guidance.”
- 5 Kiesewetter’s “misguidance argument” also seems to fall into the category of arguments that only make minimal assumptions about action guidance (“How Reasons Are Sensitive to Available Evidence”).

scribing *B* will actually cure the patient, which—if there is no epistemic filter that reason-giving facts must pass—we can plausibly assume to be a decisive reason to prescribe *B*. Perspectivists, however, will say that Jill ought to prescribe *A*, because, on their view, what one ought to do depends on one's epistemic perspective on the facts. Since Jill's available evidence decisively suggests that only *A* will cure the patient, it is this pill that she ought to prescribe according to perspectivism.

As I have indicated, there is a question about what exactly it means for a normative notion like “ought” to depend on one's epistemic perspective.⁶ Since my argument against objectivism does not depend on this question, I will not say much about it here. To simplify the presentation and without further argument, I will adopt Benjamin Kiesewetter's elaborate version, which has been particularly influential in the recent literature and whose central thesis is the following:⁷

Evidence-Relative Perspectivism: *S* ought to ϕ if and only if *S*'s available reasons decisively favor ϕ -ing.

For our purposes, it will suffice to work with an intuitive notion of “available reasons.”⁸ In *Doctor*, Jill has a body of evidence available to her—such as a review of recent articles from medical journals and the opinion of her well-respected colleagues—that gives her strong reasons in favor of prescribing *A*. Furthermore, all facts that might count against prescribing *A*, including the fact that it will kill the patient, are epistemically unavailable to her in any intuitive sense: by assumption, she does not know that these facts obtain, nor is she in a position to know or even justifiably believe them. On a perspectivist reading, any plausible normative theory will thus entail that Jill's reasons to prescribe *A* are overall decisive and so she ought to prescribe *A*.

We have seen that objectivism and perspectivism entail conflicting verdicts about what Jill ought to do. Before I discuss the relative plausibility of these views, however, let me try to preempt (though perhaps not decisively) a suspicion that might have occurred to the reader by now: that the dispute between objectivists and perspectivists is merely verbal.

In particular, we could follow the lead of what we can call *disambiguationist* views, which recommend distinguishing between two different notions of

6 See, e.g., Zimmermann, *Ignorance and Moral Obligation*; Kiesewetter, “What Kind of Perspectivism?”

7 Kiesewetter, “‘Ought’ and the Perspective of the Agent,” “How Reasons Are Sensitive to Available Evidence,” *The Normativity of Rationality*, and “What Kind of Perspectivism?”

8 For further discussion, cf. Kiesewetter, “How Reasons Are Sensitive to Available Evidence” and “What Kind of Perspectivism?”

“ought.” On such a view, Jill ought *objectively* to prescribe *B*, but ought *subjectively* to prescribe *A*. With such a distinction at hand, we might then say that objectivists and perspectivists do not have a substantive disagreement. They would both be right about Jill in different senses.

Like most other parties to the debate, however, I do not think that the problem can be solved so easily.⁹ The reason for this lies in the nature of the deliberative “ought,” which is the kind of “ought” that we are interested in here. A judgment of the form “I ought to ϕ ” is supposed to answer the deliberative question, “What ought I to do?” in a way that concludes my deliberation and allows me to make a rational decision on the basis of my deliberation. Importantly, an “ought” judgment can do this only if it employs a univocal notion of “ought.” For I might know in a given situation that I ought *subjectively* to prescribe *A* and that I ought *objectively* not to prescribe *A*.¹⁰ But in order to conclude my deliberation and make a decision on this basis, I still need to know what I ought to do, all things considered. There is no point in disambiguating, again, between a sense in which I ought all things considered *subjectively* to prescribe *A* and a sense in which I ought all things considered *objectively* not to prescribe *A*, for the same question will arise all over again. Unless there is, at some point, a univocal answer to the question of what I ought to do *period*, I cannot conclude my deliberation in the way that is necessary for me to make a rational decision on the basis of such deliberation. What objectivists and perspectivists have a substantive disagreement about is whether what one ought to do in the univocal deliberative sense is perspective dependent or not.¹¹

- 9 People who share this assessment include, e.g., Graham, “In Defense of Objectivism about Moral Obligation”; Kiesewetter, “How Reasons Are Sensitive to Available Evidence”; Jackson, “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection”; Kolodny and MacFarlane, “Ifs and Oughts”; Lord, “Acting for the Right Reasons, Abilities, and Obligation”; Piller, “Ewing’s Problem”; and Way and Whiting, “Perspectivism and the Argument from Guidance.”
- 10 For example, suppose a doctor has three pills available to her: pill *A* cures the patient, pill *B* significantly improves the patient’s condition (without curing her completely), and pill *C* kills the patient. The doctor knows about *B*’s effects, but she does not know which of *A* and *C* cures and kills the patient, respectively. (This is the famous *Jackson Case* in Jackson, “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection.”) In this example, the doctor might well know that she ought *subjectively* to prescribe *B* but that she ought *objectively* not to do this (since she knows that she ought *objectively* to prescribe either *A* or *C*).
- 11 For a helpful, more detailed discussion of the problems with disambiguationism, see, e.g., Henning, “Normative Reasons Contextualism.” There is another family of views that one could take about the perspective dependence of normative notions like “ought” or “reason,” which is similar in spirit to disambiguationism but more sophisticated in its development. According to these broadly contextualist or relativist views, there is indeed a univocal notion of “ought” or “reason,” but a statement containing these notions either expresses differ-

2. ACTION GUIDANCE

This section has two aims. First, I set out two central assumptions of my argument, one about the possibility of action guidance and the other about the relationship between objectivism in the practical and the epistemic domains. Second, I offer a substantive but still sufficiently generic account of what it is for a normative theory to be correctly action guiding.

2.1. Preliminaries

Let me begin with two important assumptions. Although I will briefly explain the motivation behind each of them, I will not defend these assumptions in any detail. My aim at this point is just to lay them out as clearly as possible.

The first assumption is one that I have already mentioned: it is the first premise of the Argument from Action Guidance, which says that it is sometimes possible for an agent to be correctly guided by a normative theory. It is important to see why this is an extremely weak, and correspondingly plausible, assumption.

I will say more about what exactly it means to be correctly guided by a normative theory later. For now, let us stick to an intuitive notion of action guidance and focus on the modal status of the assumption. Intuitively, being correctly guided by a normative theory amounts to something like this: an agent does what a normative theory tells her to do because she correctly recognizes that this is what the theory tells her to do. (Note that this notion of correct guidance is an essentially relative or procedural one: if your ϕ -ing is correctly guided by some theory, this does not mean that your ϕ -ing is correct, only that it is correct according to that theory, or that you have correctly applied the theory in arriving at this verdict.)

Especially if you happen to be a professional philosopher, it is not unlikely that you are familiar with this phenomenon. It is not unlikely, that is, that at some point in your life, you have asked yourself which course of action a particular normative theory would recommend in your circumstances, and perhaps, given that you also believed the theory in question, you actually followed that

ent propositions in different contexts or expresses the same proposition whose truth value varies according to context (for an expressivist variation on this theme, see, e.g., Björnsson and Finlay, "Metaethical Contextualism Defended"; Henning, "Normative Reasons Contextualism"; Kolodny and MacFarlane, "Ifs and Oughts"; Pittard and Worsnip, "Metanormative Contextualism and Normative Uncertainty"; and Schroeder, "Getting Perspective on Objective Reasons"). However, a detailed discussion of such views is well beyond the scope of this paper. My aim is only to raise a new challenge for objectivism, which is one that perspectivism turns out to be able to meet. But I want to remain neutral about the merits of perspectivism vis-à-vis these other possible views.

recommendation. If this is what you have done, and you were right about what the theory recommended, then you have been correctly guided by a normative theory in the sense that is relevant for my argument (perhaps setting aside contrived cases, e.g., those involving deviant causal chains).

More importantly, even if you have never actually been guided by a normative theory in this sense, this does not make it any less plausible that one *could have been* guided in this sense. All that needs to be the case for that claim to be true is for it to be possible for someone, under some circumstances, to correctly apply a normative theory in practice. This is all that the first premise requires and it is, as I have said, very plausible.

This is in large part because the modal status of the premise is so weak. For all that it says, it might be the case that whenever an agent faces a decision under ignorance, it will no longer be possible for her to be correctly guided by a normative theory. This is consistent with my assumption about guidance, as long as there remain some cases where an agent *is* correctly guided by a normative theory. In general, the premise does not require that, in any possible situation, there is at least one normative theory that could correctly guide the agent, let alone that there is a single normative theory that could always do that. These are the sorts of stronger guidance assumptions that are likely to beg the question against objectivists. Since my argument makes a much weaker assumption, it cannot be rejected on these grounds.

In fact, some objectivists, such as Peter Graham, explicitly embrace the idea that normative theories can correctly guide action.¹² Graham even accepts this possibility in cases of ignorance, and he explains in detail how a normative theory could correctly guide a morally conscientious agent in such cases, even if objectivism is true. Similarly, Way and Whiting argue on behalf of objectivism that this view is perfectly consistent with the idea that normative theories sometimes provide correct action guidance.¹³ This is further evidence that the weak guidance premise is independently plausible and does not beg the question against objectivists. I will hence assume it in what follows.¹⁴

12 Graham, "In Defense of Objectivism about Moral Obligation."

13 Way and Whiting, "Perspectivism and the Argument from Guidance."

14 Another consideration is that one of the main motivations for objectivism—that it is the best account of our practice of advice (see, e.g., Graham, "In Defense of Objectivism about Moral Obligation")—fits quite badly with denying that the normative is possibly action guiding. This is because it is hard to see why advice would be an important normative phenomenon if it were not at least sometimes possible for normative claims to correctly guide action. Why would people ask for advice if they did not think that responses like "You ought to ϕ " could correctly guide action, and why would people respond to a request for advice if they did not think that their advice could correctly guide others? The fact that we often seek

The second assumption that I will make is that the truth of objectivism about the practical “ought” entails the truth of objectivism about the epistemic “ought.”¹⁵ This assumption is motivated on several grounds. First, whatever substantive differences between the practical and the epistemic “ought” there might be, they still seem to be instances of the same distinctive notion of all-things-considered normative assessment. Assuming that this general notion of “ought” does not behave radically differently in the two domains should thus be the default position, unless there is some positive reason for denying it. But it is far from obvious what this positive reason could be.

Second, we must keep in mind that some of the best arguments for practical objectivism are based on considerations about the practice of advice. According to one such argument, the fact that true advice about what someone ought to do can be based on the facts rather than the advisee’s evidence suggests that what the advisee ought to do must itself depend on the facts rather than her evidence.¹⁶ But since it seems that true advice about what someone ought to *believe* can similarly be based on the facts rather than the advisee’s evidence, the very same kind of argument would support epistemic objectivism. One of the central arguments in favor of practical objectivism thus generates an internal pressure to accept epistemic objectivism as well.

Third, the combination of practical objectivism and epistemic perspectivism yields conflicts with the plausible and widely acknowledged claim that it is irrational to believe that one ought to ϕ without intending to ϕ .¹⁷ If objectivism were true only in the practical domain, then in cases like *Doctor* it would both be true that Jill ought to believe that she ought to prescribe pill A (since this is, given her evidence, most likely to be the cure) and that she ought not to intend to prescribe pill A (since, we can stipulate, this would prevent her from prescrib-

and give advice, a fact that objectivists regularly emphasize, suggests that we do accept that normative claims, and so the normative theories constituted by them, can correctly guide action.

15 This assumption is widely shared in the literature. It is implicit in Gibbons (*The Norm of Belief*, 1–17), who takes the same considerations that motivate practical objectivism to motivate epistemic objectivism as well. Thomson thinks that objectivism has to be true in both the practical and epistemic domains (*Normativity*, chs. 11 and 12). The assumption that both domains ought to be treated analogously is also important in McHugh and Way, “Objectivism and Perspectivism about the Epistemic Ought,” 139–40; and Littlejohn, *Justification and the Truth-Connection* and “The Unity of Reason.”

16 Cf. Graham, “In Defense of Objectivism about Moral Obligation,” 91.

17 Cf. Scanlon, *What We Owe to Each Other*, secs. 2.3, 2.4; Broome, *Rationality through Reasoning*, sec. 9.5; Kolodny, “How Does Coherence Matter?”; Kiesewetter, *The Normativity of Rationality*; Way, “The Symmetry of Rational Requirements”; Lord, “The Real Symmetry Problem(s) for Wide-Scope Accounts of Rationality.”

ing pill *B*, which is what she ought to do according to objectivism). This means, however, that if Jill conforms to both “oughts,” she will necessarily be irrational. But it is implausible that in cases like *Doctor*, and all other cases where people have misleading evidence about what they ought to do, conforming to all the “oughts” to which one is subject is a way of ensuring one’s own irrationality.¹⁸ The fact that a “hybrid” view about the perspective-dependence of “ought” entails this result is a strong reason to reject it.

2.2. A Generic Account of Action Guidance

With the two guiding assumptions of my argument on the table, let me now turn to the question of what exactly it means for a normative theory to correctly guide action. I will state my proposed analysis before I discuss each of its specific conditions:

Action Guidance: A normative theory *T* correctly guides an agent *S* toward ϕ -ing if and only if

- (A) *S* ϕ -s on the basis of correct practical reasoning from *T* toward ϕ -ing, with *A* being satisfied only if
- (B1) for some *R* that is a decisive reason to ϕ according to *T*, *S* believes that *R* is a decisive reason to ϕ , and
- (B2) *S* believes that *R* is a decisive reason to ϕ for the reason that *E*, with *E* being the right reason to believe this according to *T*.

In this analysis, *R* is any practical reason, or set of practical reasons, that decisively favors ϕ -ing according to some theory *T*. *E* is the right reason, relative to that theory, to believe that *R* is a decisive (practical) reason to ϕ . (The qualifications are important: the requirement is not that *R* and *E* really *are* reasons for acting and believing, only that *T* says so. I will come back to this issue soon.)

Turning now to the specific conditions asserted by the account, let us first ask what justifies *A*. The main idea is that a reasonable account of action guidance must explain the difference between being guided by a theory and merely acting in conformance with it. Consider the following case:

Utilitarian Demon: Since Jim believes in utilitarianism, he believes that he ought to donate a significant portion of his wealth to an efficient charity. But since he is akratic, he cannot get himself to donate. Unbeknownst

18 See Littlejohn, *Justification and the Truth-Connection*, 235, for essentially the same argument. Note that the argument does not assume that epistemic and practical “oughts” agglomerate (i.e., that “*S* ought to believe *X*” and “*S* ought to do *Y*” entails “*S* ought to believe *X* and do *Y*”). It only assumes that, if you conform to each individual “ought” that applies to you, you will end up with an irrational combination of attitudes.

to him, there is a demon who enjoys making akratic utilitarians perform what utilitarianism requires. The demon tinkers with Jim's brain and, consequently, Jim intentionally donates a significant portion of his wealth to an efficient charity.

Let us assume that Jim acts in accordance with utilitarianism, because utilitarianism really requires what Jim (and the demon) thinks it requires. Still, Jim is not correctly guided by utilitarianism. This is so although Jim believes in utilitarianism and intentionally does what utilitarianism requires, and there is a clear connection between his belief and his action.

However, the connection is of the wrong kind. Although Jim's doing what utilitarianism requires is causally connected to his belief, the move from belief to action is not *guided* by utilitarianism. For an agent to be correctly guided by a theory, a mere causal connection is not enough. Instead, I suggest, the connection must be one of correct practical reasoning *from* that theory *toward* the action required by it. In cases like *Utilitarian Demon*, the agent does not act on the basis of correct practical reasoning, and so is not correctly guided by the normative theory in question. This is why we should include A as a necessary condition for correct action guidance.

Condition A raises the following question: What makes practical reasoning correct? Obviously, I cannot provide a substantive answer here. For our purposes, it will be enough to rely on a formal answer: correct practical reasoning just is reasoning that conforms to the correct standards of practical reasoning, whatever they are. In this context, what is more important than the substance of these principles is what notion of correctness is relevant here.

Again, this is a form of procedural correctness. Roughly speaking, practical reasoning is correct in a procedural sense if it transforms correct inputs into correct outputs. The idea is not that if correct reasoning issues in a deliberative conclusion of the form "I ought to ϕ " it will be true that I ought to ϕ . Rather, the point is that, if the premises of the reasoning are true *and* one reasons correctly from this basis, one will be guaranteed to arrive at a true deliberative conclusion. For example, if one engages in (procedurally) correct reasoning from Kantian premises, it will be guaranteed that, if these Kantian premises are true, so are the deliberative conclusions arrived at in such reasoning.

To sum up, being correctly guided by some normative theory *T* requires that the transition from believing the theory toward action is mediated by correct practical reasoning, in a procedural sense of "correct." But what exactly does this involve? In order to provide a (partial) answer to this question, let me now turn to conditions B₁ and B₂, which I take to be necessary conditions for it to be true

that someone acts on the basis of correct practical reasoning from a given normative theory.

According to B₁, this is the case only if the agent believes *R* to be a decisive reason to ϕ , for some practical reason, or set of practical reasons, *R*, that decisively favors ϕ -ing according to the normative theory in question. Why is this a necessary condition for correct action guidance?

The motivation is a broadly Kantian one (although it is not just Kantians who will find the underlying picture attractive). As Alison Hills expresses the thought in a related discussion of moral worth and what it is to act for moral reasons:

This “Kantian” theory says that to act for moral reasons requires that when you act, *you are aware of the moral reasons in favour of your action, you are aware of their normative force* (e.g. that the action is required, permitted, or favoured by them), and you decide to act and put that decision into action on the basis of your awareness of that force, responding appropriately, that is, rationally to it.¹⁹

According to the Kantian theory, acting for a reason requires one’s awareness of these reasons *as reasons*. If we apply this thought to the question of what it is to be correctly guided by a normative theory, we should similarly say that it requires one’s awareness (as reasons) of those normative reasons that the theory identifies as normative reasons.

To see how plausible B₁ is, we can also consider what it would mean to deny it. For suppose that, according to *T*, only *R* is a decisive reason to ϕ , but that the agent ϕ -s without taking *R* to be a decisive reason to ϕ . Whatever her ϕ -ing is based on, this will *not* be something that the theory acknowledges to require ϕ -ing. If the agent ϕ -s in such a case, then she either ϕ -s for no reason at all, or she ϕ -s for some other reason that does not decisively favor ϕ -ing according to *T*. In neither case is the agent appropriately described as ϕ -ing on the basis of *correct* reasoning from *T*, a theory that says that only *R* requires ϕ -ing. The same is true if according to *T* there are multiple decisive reasons for ϕ -ing, but the agent does not take any of them to require ϕ -ing. Therefore, we should incorporate B₁ as a necessary condition into our account of correct action guidance.

According to B₂, *S* ϕ -s on the basis of correct practical reasoning from *T* toward ϕ -ing only if *S* believes for *the right reason E* that *R* decisively favors ϕ -ing (where, again, what makes a reason the “right” reason to believe something is

19 Hills, “What Does It Take to Act for Moral Reasons?” 247, emphasis added. Hills herself later rejects this “rational guidance” account, because she thinks that it is overly intellectualized. In section 2.3, I explain why my account of action guidance does not fall prey to this objection.

determined by the theory in question). In contrast to R , which is a reason for action (relative to T), E is a reason for *belief* (relative to T). Again, it is important to stress this last qualification: E may not *really* be a reason to believe that R decisively favors ϕ -ing. All that is required by B2 is that E is what T *claims to be* the right reason to believe this.

To see why this is necessary, suppose that the correct analysis of action guidance did not include B2, and consider the following example. You borrow money from your friend, promising to pay it back in due time. After a while, it occurs to you that keeping the money would actually be quite a good bargain. So you ask yourself: "Ought I really to pay it back?" When you remember your promise, you wonder whether that is not a good reason to return the money. Now, there are a variety of reasons for which you could believe that having promised something is a decisive reason to do it. You might think that your promise generates a decisive reason because breaking the promise decreases overall utility, or because this is prohibited by God's commands, or because it violates a particular code of honor, or because it is ruled out by the categorical imperative. And which of these thoughts you think must determine whether you can be correctly described as being guided by this or that particular normative theory.

Suppose you think that having promised to return the money is a decisive reason to do it because God commands that promises be kept. Then your returning the money, if that is what you end up doing, is not correctly guided by Kantianism, even if Kantianism *also* says that having promised to return the money is a decisive reason to do it. Although Kantianism (let us suppose) says that your promise generates a decisive reason, and you believe that your promise generates a decisive reason, you do not believe this because you take promise breaking to be ruled out by the categorical imperative, or because of some other recognizably Kantian explanation. You believe that your promise generates reasons because God has commanded that promises be kept. And so while your reasoning might be correct by a divine command theory standard, it is not correct by a Kantian standard.

Therefore, being guided by Kantianism does not just require an agent to believe that her promise decisively favors performing the promised action. It also requires that she believes *this* for the reason that explains, according to Kantianism, why her promise generates a decisive reason. This is why we need B2 on top of B1.

3. OBJECTIONS

Since my account of correct action guidance is a central element in the argument

to come, it will be worthwhile to address a number of possible objections to it. In particular, I want to anticipate certain worries about B₁ and B₂, two conditions that carry much weight in the following argument.²⁰

3.1. *Is the Account Overly Intellectualized?*

The first objection presses the worry that B₁ yields an implausibly intellectualized account of action guidance. Are there not ways of being guided by a normative theory, one might ask, that do not require the agent to have the concept of a normative reason, let alone explicit beliefs about such things?

Hills discusses a similar objection to the Kantian account of acting for moral reasons mentioned above. In her view, it is sometimes perfectly felicitous to describe an agent as having acted for the right moral reasons, even if she was not aware of what these reasons were and hence did not believe them to favor the action. For example, before the concept of sexual harassment became widely known, a woman might have quit her job for the right moral reason that she was harassed, even if she did not believe that she was harassed or that being harassed is a moral reason to quit one's job.²¹ Similarly, one might think that agents could be guided by a normative theory "implicitly" or "subconsciously," without an explicit awareness of how the considerations that are reasons according to the theory in question favor a certain action.

In response, I want to point out that nothing in my argument commits me to denying this. For all that I say, there might well be such a thing as implicit or subconscious guidance by a theory, of which the above account is not the correct one. Still, even if that is the case, there surely is such a thing as being *explicitly* guided by a normative theory, which does involve practical reasoning and normative beliefs. It is this phenomenon that I seek to give an account of and that plays a role in the subsequent argument. It is problematic enough if objectivism entails that explicit action guidance in this sense is impossible, for surely it is possible to sometimes act on the basis of explicit guidance by a normative theory. For this reason, the charge of being overly intellectualized does not undermine my account of action guidance, even if it succeeds as an objection against parallel accounts of acting for the right reasons.

3.2. *Normative Theories with Too Little Explanatory Structure*

The second objection claims that B₂ is not a necessary condition for correct action guidance. In some cases, it seems that an agent is correctly guided by a nor-

20 I owe the following objections, including the purported counterexamples in sections 3.2 and 3.3, to two anonymous referees.

21 Cf. Hills, "What Does It Take to Act for Moral Reasons?" 253.

mative theory, although there is no reason E for which she could rightly believe (relative to that theory) some other reason R to decisively favor ϕ -ing. For example, consider a normative theory that says both that promises generate reasons *and* that nothing explains why this is so. In that case, there seems to be no E for which one could rightly believe (relative to that theory) that promises generate reasons. It would then follow from my account that one could not be correctly guided by that theory, which might seem like the wrong result.

In response, note first that, in my proposed analysis, R is not just any reason, but a decisive one. So the theory we are imagining says, not just that nothing explains why promises generate reasons, but also why, in a given situation, the promissory reasons are decisive. It is not clear to me that this is properly called a “theory.” If you think that it is simply, inexplicably true that you ought to keep your promise in a given situation, then perhaps the most plausible thing to say is that there is nothing that you have a *theory* about. (In fact, the envisaged view sounds reminiscent of W.D. Ross’s moral pluralism, which is notoriously criticized for being insufficiently action guiding, precisely because it lacks theoretical structure.)²²

But even if we grant for the sake of the argument that this is a genuine normative theory, we can still account for this complication by being liberal about what can count as E : the right reason to believe that promises generate decisive reasons, relative to different theories, might be either a substantive reason, e.g., that breaking promises violates the categorical imperative, or alternatively something like the brute fact that promises simply *do* create decisive reasons. Unless one believes that promises simply *do* create decisive reasons, the analysis then predicts—correctly, in my view—that one is not correctly guided by this sort of theory.²³

3.3. Guidance De Re and De Dicto

Let me now turn to the final objection, which puts pressure on B_1 by appealing

- 22 Ross, *The Right and the Good*. Note that I am not endorsing this as an *objection* to Rossian pluralism. I just mention it to support the claim that many people seem to think that Rossian pluralism has too little structure to count as a genuine normative theory, whether or not that is a reason to reject it.
- 23 One must also keep in mind that B_2 is merely a necessary condition. Several different theories might all claim that promises simply *do* create decisive reasons. If you keep a promise on the basis of such a thought, conditions B_1 and B_2 are silent on which of these theories has correctly guided you. Since they are not sufficient conditions, my account does not implausibly entail that, whenever you keep your promise on the basis of such a thought, you automatically count as being guided by any normative theory according to which the normative force of promises has no further explanation.

to cases where agents seem to be correctly guided by disjunctive considerations. To illustrate, consider the following example:

Disjunctive: Yesterday, Jane borrowed Jack's pen. Today, she asks herself whether she ought to return it. Jane accepts normative theory *D* according to which one ought to ϕ if and only if either (i) ϕ -ing keeps a promise or (ii) ϕ -ing maximizes overall utility. She also believes that returning Jack's pen either keeps a promise or maximizes overall utility (but not both), and decides to return the pen on the basis of accepting this disjunction. As a matter of fact, Jane has indeed promised to return Jack's pen, but returning it does *not* maximize overall utility.²⁴

Could Jane count as having been correctly guided by *D*? Intuitively, yes. But according to *Action Guidance*, the answer seems to be no. This is because, according to *B1*, an agent is correctly guided by a normative theory *T* only if she correctly believes, for some *R* that is a decisive reason according to *T*, that it decisively favors her action. But in *Disjunctive*, it seems that there is no such *R*.

To see why, consider three possible candidates: the true proposition (a) that Jane has promised to return the pen; the false proposition (b) that returning the pen maximizes overall utility; the true proposition (c) that returning the pen either keeps a promise or maximizes overall utility.

It cannot be (a) or (b), because Jane believes neither of them. The only remaining candidate is thus (c). The problem is that, as the case is described, the disjunction (c) is *not* a reason according to *D* (whatever Jane believes about it).²⁵ According to *D*, only two considerations could be a reason for returning the pen: (i) that doing so keeps a promise, or (ii) that it maximizes overall utility (but not their disjunction). And still, it both seems that Jane could be properly guided by the disjunction and that there is a natural way of describing the resulting action as being correctly guided by *D*. If that is right, *B1* cannot be a necessary condition for correct action guidance.

What is the right way to reply to this objection? I think a careful description of the case points toward a solution. Why is it intuitively plausible in the first place to say that Jane has been correctly guided by *D*? It is a crucial element of the case that *D*'s criterion of rightness is a disjunctive one. It is for this very

24 Admittedly, this is a contrived case. Perhaps the easiest way to imagine it is by stipulating that Jane loses her memory after making the promise and that a reliable informant later tells her that returning the pen either keeps her promise or maximizes overall utility (but not both).

25 We can leave it open whether *D* is right to exclude disjunctive considerations from being reasons. What matters is only that it does. The mere fact that a normative theory is mistaken does not entail that one cannot be correctly guided by it, in the procedural sense of "correct."

reason that Jane does not *need* to know which of the specific *R*s that would be decisive reasons according to *D* actually obtains. She is still in a position to know *that* something requires ϕ -ing according to *D*, even if she cannot know *which* fact makes this the case. And that puts her in a position to correctly apply *D* as a guide to action.

What does this mean for the status of *Action Guidance*? I think that reflection on the example shows that there are in fact two interestingly different ways in which B_1 can be satisfied. On the first reading, the agent must correctly believe, for some specific consideration *R*, that it is a decisive reason according to *T*, in the sense of being able to actually say what that consideration is (e.g., that the action keeps a promise). Call this the *de re* reading. On the second reading, the agent must correctly believe that *some* consideration *R* is a decisive reason according to *T*, but she need not be able to specify what exactly this consideration is. In other words, she only needs to correctly believe that some consideration, *whatever it is*, is a decisive reason according to *T*. Call this the *de dicto* reading.

Once we make this distinction, we can see that there are two different ways for B_1 to be true. In ordinary cases, the agent will be guided in the *de re* sense by her correct beliefs about which specific considerations are decisive reasons according to the theory in question. However, in cases like *Disjunctive* where guidance *de re* is unavailable, the agent can still be guided in the *de dicto* sense by her correct belief *that* some consideration is a decisive reason according to that theory, even if she is unable to specify which it is. But since correct guidance *de dicto* is correct guidance all the same, B_1 properly understood actually vindicates the intuitive verdict that Jane could be correctly guided by *D*.²⁶ We must only keep in mind that it also has a *de dicto* reading.

This completes my discussion of potential objections to *Action Guidance*. In

26 Perhaps you reject the claim that guidance *de dicto* is a genuine case of correct guidance, partly on the basis of cases like the following: Jack accepts Kantianism; a reliable informant correctly tells him that (but not why) Kantianism requires ϕ -ing; Jack ϕ -s on the basis of such testimony. Has Jack been correctly guided by Kantianism? I myself do not have a very clear intuition about this. I am inclined to say that he has (albeit perhaps not to the same degree as in the “pure” cases), since there is after all a recognizable sense in which his ϕ -ing is essentially due to his acceptance of Kantianism and a correct belief about what it requires. But the point is this: *if* you think that an agent has to know *which* specific consideration decisively favors ϕ -ing according to a theory in order to count as being correctly guided by it, then cases like *Disjunctive* should not strike you as convincing counterexamples to begin with. If, that is, only guidance *de re* is correct guidance, then Jane’s clearly is not a case of someone being correctly guided by *D*, since her returning the pen is not guided by an appreciation of a specific consideration that favors this action according to *D*. On this view, there is no problem for *Action Guidance* in the first place, as long as one insists that only the *de re* reading of B_1 is legitimate.

the next section, I will use this account to develop my main argument against objectivism.

4. WHY OBJECTIVISM RULES OUT ACTION GUIDANCE

To repeat, being correctly guided by a normative theory is a matter of moving from that theory toward an action required by the theory via correct practical reasoning. This reasoning is correct only if the agent correctly believes that R decisively favors her action, for some R that does this according to the theory, and she believes *this* for the reason E , which is the right reason to believe this according to the theory. This section argues on the basis of this account that, if objectivism is true, it is impossible for a normative theory to correctly guide action in this sense.

4.1. *The Infection Argument*

Let me start with the main idea: objectivists about practical normativity are committed to objectivism about epistemic normativity. But since epistemic objectivism makes it impossible to believe normative facts for the right reason, it thereby *infects* the action-guiding capacity of normative theories. Here is a summary of the argument:

The Infection Argument

1. Epistemic objectivism implies that it is impossible to believe normative propositions for the right reason.
2. One is correctly guided by a normative theory only if it is possible to believe normative propositions for the right reason.
3. Therefore, epistemic objectivism implies that one is never correctly guided by a normative theory (from 1 and 2).
4. Furthermore, practical objectivism implies epistemic objectivism.
5. Therefore, practical objectivism implies that one is never correctly guided by a normative theory (from 3 and 4).

I have already defended 2 above. Even if someone believes that R decisively favors ϕ -ing, she is not correctly guided by that theory if she does not believe *this* for the reason that explains, according to the theory in question, why R decisively favors ϕ -ing. Furthermore, 4 is among the guiding assumptions that I have briefly defended above. Hence, only 1, which says that epistemic objectivism is inconsistent with believing normative propositions for the right reason, needs further defense. The rest of this section embarks on this task.

What exactly is epistemic objectivism? Roughly speaking, it is the view that

what one ought to believe depends on all the facts, rather than on one's epistemic perspective.²⁷ It contrasts with epistemic perspectivism, which says that what one ought to believe depends only on epistemically available facts, such as the evidence that is available to the agent during doxastic deliberation.

In order to characterize epistemic objectivism more precisely, we should follow suggestions by Daniel Whiting, Conor McHugh, and Jonathan Way and deny that, according to epistemic objectivism, *S* ought to believe *p* if and only if *p* is true.²⁸ We should not, that is, ascribe to epistemic objectivists the view that one ought to believe just any old truth, no matter how trivial or uninteresting. On a more careful characterization, epistemic objectivism claims that it is *permissible* to believe *p* if and only if *p* is true.²⁹ This gives epistemic objectivists all that they should want: it is not the case that one ought to believe just any old truth, but what it is permissible to believe is still fully dependent on all the facts rather than one's epistemic perspective.

The remainder of the section proceeds in three steps. First, I argue that if one accepts epistemic objectivism one should accept that *p* is a reason to believe *p*. Second, I argue that epistemic objectivists who accept that *p* is a reason to believe *p* should also accept that *only p*—as opposed to independent evidence bearing on whether *p*—is a reason to believe *p*. Third, I explain why, if only *p* is a reason to believe *p*, for some normative proposition *p*, it is impossible to believe *p* for the right reason.

4.2. *If Epistemic Objectivism Is True, Then p Is a Reason to Believe p*

Let us start with the first commitment of epistemic objectivism. Why does epistemic objectivism entail that *p* is a reason to believe *p*?³⁰ First, recall that, accord-

27 For versions of epistemic objectivism, see, e.g., Boghossian, "Is Meaning Normative?"; Gibbard, "Truth and Correct Belief"; Littlejohn, *Justification and the Truth-Connection*; Schroeder, "Knowledge Is Belief for Sufficient (Objective and Subjective) Reasons"; Shah, "How Truth Governs Belief"; Shah and Velleman, "Doxastic Deliberation"; Wedgwood, "The Aim of Belief"; and Whiting, "Should I Believe the Truth?"

28 Whiting, "Should I Believe the Truth?"; McHugh and Way, "Objectivism and Perspectivism about the Epistemic Ought."

29 Some people might find it unnatural to speak of what it is "permissible" to believe, as opposed to which beliefs are justified or warranted. I myself do not find this expression problematic, as long as one keeps in mind that responding to what it is permissible to believe need not be thought of in the same way as responding to what it is permissible to do. (Typically, one does not first register a permission to believe *p* and then go on to believe it, in the same way in which one might first register a permission to do something and then go on to do it.) Furthermore, speaking of epistemic permissions is common practice in the debate that I am engaging in here, and I will rely on it in what follows.

30 This commitment (or something close to it) seems to be explicit in McHugh and Way, "Ob-

ing to epistemic objectivism, the normative status of believing p depends on p 's truth. But if that is so, what could speak more clearly in favor of believing p than the fact that p is true? If the agent's epistemic perspective does not affect the permissibility of a belief, and what it is permissible to believe depends instead on all the facts, then why exclude the fact that p ?

Let us say that if the conditional probability of q given p is greater than the unconditional probability of q , then p is a probabilifier with respect to q . Now suppose epistemic objectivists deny that p is a reason to believe p . On this view, if there are any reasons to believe p at all, then probabilifiers q , r , s , etc. are reasons to believe p , while p is not a reason, although p is a probabilifier just like q , r , s , etc. (It does not fail to be a probabilifier just because it raises p 's probability to 1 rather than, say, 0.9.) Indeed, since it maximally increases the probability of p , it is, if anything, the *paradigm case* of a probabilifier with respect to p , in which case p seems to be a reason to believe p just like the other probabilifiers. Therefore, it seems that if objectivism is true, the fact that the candidate in fact won the election is a reason to believe that she won the election.³¹

Of course, perspectivists have grounds for denying that p is a reason to believe p . They believe that reasons must be epistemically available facts that one can rationally appeal to in deliberation. But cases of doxastic deliberation about p are precisely cases where the agent is *about to decide* whether p is true, which explains why *at that time* p is not among the available facts that one can rationally appeal to. Hence it is not a reason as far as perspectivism is concerned. But since objectivists deny that reasons must be epistemically available in such a sense, they cannot reject that p is a reason on these grounds. Unless one follows perspectivism to accept such restrictions on what can count as a reason, it just

jectivism and Perspectivism about the Epistemic Ought," 133, and Schroeder, "Knowledge Is Belief for Sufficient (Objective and Subjective) Reasons," sec. 2, who assume on behalf of objectivism that p 's falsity is a reason *not* to believe p —if we make this assumption, it also seems reasonable to accept that p 's truth is a reason *to* believe p . Littlejohn, himself an objectivist, seems to accept this for normative facts, which we are concerned with here: "Suppose that it is a fact that A ought to ϕ . If so, this fact should be part of the evidence there is. If it is evidence for anything, it should be evidence for itself" ("Do Reasons and Evidence Share the Same Residence?" 4). Presumably, he would then say that, since evidence for p is a reason to believe p , p is a reason to believe p .

- 31 Cf. Matt Weiner: "What we should believe with all the facts at our disposal and unlimited mental power to process them ... are all and only the true propositions. If we could help ourselves to all the facts, then we could help ourselves to the facts about whether any given proposition is true, and that is enough to determine that we should believe it" ("The Spectra of Epistemic Norms," 205–6). As Weiner says, if one were aware of p 's truth, this would be "enough to determine" that one should believe p . In such a case, p 's truth would be a (decisive) reason to believe p .

seems *ad hoc* to exclude p from the class of probabilifiers that are reasons to believe p .

Second, denying that p is a reason to believe p makes the intuitive motivation for epistemic objectivism look quite sparse. Objectivism would then say that probabilifiers are reasons to believe some proposition, as long as they are different from that very proposition. Since probabilifiers are sometimes *non-veridical*, probabilifying false propositions, there will be reasons to believe false propositions even on objectivism so understood.³² But if objectivists already allow that there can be reasons to believe false propositions, it just becomes odd to maintain that these reasons are *necessarily* overridden, for any proposition and under all circumstances. If these are *reasons* with genuine normative force, then would not it have to be possible, at least under certain circumstances, for them to permit a belief that is false but extremely well supported by independent evidence?

In general, it does not look like an intuitively attractive combination of views to think *both* that independent evidence for some falsehood p can provide genuine reasons to believe p *and* that it is necessarily impermissible to believe a falsehood, even when the independent evidence in support of it is incredibly strong. Objectivism would seem much better motivated if it claimed that, since believing as one ought to is a matter of believing truly (rather than having beliefs that are *likely* to be true), only the fact that p is true (rather than that it is *likely* to be true) is a reason to believe p . (We will see shortly why it is “only” p ’s truth.)

Third, it is very plausible to assume that reasons are considerations that can make something permissible. In the practical case, for example, reasons to spend time with my friends tonight rather than work an extra shift in the office are considerations that, if sufficiently strong, can make it permissible to spend time with my friends tonight. Return now to the epistemic case: if p itself were not a reason to believe p , it turns out that this natural view is mistaken. If epistemic objectivism is true, then p ’s truth alone makes it permissible to believe p . But it is not, on the present suggestion, a reason to believe p . At the same time, the considerations that *are* supposed to be reasons to believe p —independent evidence for p —are not the kinds of considerations that make believing p permissible. Either p is true, in which case it is permissible to believe p whatever the evidence says, or p is false, in which case it is impermissible to believe p whatever the evidence says. Whatever independent evidence there is, it does not make any difference to the permissibility of a belief. Hence, if p were not a reason to believe p , this would implausibly entail that reasons for belief are *not* the kinds of consideration

32 Technically, it could be that only veridical probabilifiers are reasons. But that view looks even more gerrymandered and without any intuitive motivation. I disregard it here.

that can make belief permissible. For these reasons, I conclude that, if epistemic objectivism is true, then p is a reason to believe p .

4.3. *If Epistemic Objectivism Is True, Then Only p Is a Reason to Believe p*

Consider now the second step and ask whether objectivists could say that both p 's truth and independent evidence bearing on whether p are reasons to believe p . The problem is that this view conflicts with at least two plausible constraints on reasons.

The first constraint is:

Impact Constraint: If R is a reason to ϕ , then R could make a difference, or contribute, to its being the case that one ought to ϕ .

I take it that there is an intuitive sense in which reasons are the kinds of things that make a difference to "ought" facts. Whenever we weigh up reasons for and against something, we do so precisely because we thereby try to track those considerations that help to make it the case that we ought (or ought not) to respond in a certain way. In the practical case, for example, it is very natural to think that something like my having promised to ϕ makes a difference, or contributes, to it being the case that I ought to ϕ (even if all things considered I ought not to ϕ).

However, as I have already suggested, independent evidence fails to be difference making in this sense. If p is true, it is *ipso facto* permissible to believe p whatever the independent evidence says; if p is false, it is *ipso facto* impermissible. If epistemic objectivism is true, independent evidence bearing on p , in contrast to p 's truth itself, does not have any influence on the permissibility of a belief.

Indeed, one can say more. A useful, maximally weak heuristic for finding out whether a consideration of a certain kind makes the right kind of difference in order for it to be a reason seems to be this: suppose first that ϕ -ing is impermissible. Then ask yourself whether it is at least conceivable that, by adding more and more considerations of that kind, there is a point at which the scale could tip and ϕ -ing becomes permissible (or even required). Now, it is clear that, if epistemic objectivism is true, independent evidence for p fails this test. If believing p is impermissible, this means that p is false. And so there will not conceivably be a point at which enough independent evidence is added to make the belief that p permissible (or even required). This confirms the intuitive verdict that, if objectivism is true, independent evidence cannot be a reason, precisely *because* it does not make any difference to all-things-considered facts about what one is permitted to believe.

The second constraint is:

Weighing Constraint: If R is a reason for or against ϕ -ing, then there is a conceivable situation where it makes sense to weigh up R and other reasons in deliberation about whether to ϕ .

Weighing up reasons is what we do whenever we work ourselves toward an “ought” conclusion on the basis of competing contributory considerations. The above constraint expresses the plausible thought that, for something to be a reason, it must at least conceivably be the kind of thing that could play the contributory role in such reasoning.

Yet, *Weighing Constraint* rules out that both p itself and independent evidence bearing on whether p could be reasons for or against believing p . When making up one’s mind about whether to believe p , it does not make sense, even conceivably, to entertain both p and independent evidence bearing on whether p in doxastic deliberation. All cases where p ’s truth is among the reasons that one weighs in deliberation are cases where the question of whether to believe p is already settled in favor of believing it. If one of the considerations that one appeals to in doxastic deliberation is that p is true, there is no point in weighing up p ’s truth against other reasons *in order to find out whether p is true*.

It is important to note that this constraint does not beg the question against objectivism.³³ One might think that, in cases like *Doctor*, it would not make sense for Jill to weigh up the fact that pill B is the cure against other reasons bearing on her decision. After all, Jill is not aware of this fact, and it might well seem that one could not sensibly appeal to facts in deliberation of which one is unaware. But since this fact is a reason according to objectivism, it would beg the question of whether *Weighing Constraint* ruled this out.

However, this is not what the constraint does. What it rules out is that something is a reason if there is no *conceivable* situation where it can sensibly be weighed up against other reasons. This is not true for the fact that B is the cure, which one might sensibly weigh up against other reasons for or against prescribing it in all sorts of circumstances, even if in Jill’s *actual* circumstances she could not sensibly appeal to this (unknown) fact. Still, with regard to p ’s truth in deliberation about whether p , this is not even conceivably the case. As Kieseewetter says in a related context, p and independent evidence bearing on whether p

could not both be reasons in the same situation. . . . This is not because weighing them is practically impossible from your point of view, but because there is no single point of view from which these two facts can sensibly be weighed against each other.³⁴

33 I am indebted to an anonymous referee for pressing me on this and the following point.

34 Kieseewetter, “How Reasons Are Sensitive to Available Evidence,” sec. 3. Note also that my

I conclude that, if epistemic objectivism is true, then only p —as opposed to independent evidence bearing on whether to believe p —is a reason to believe p .

4.4. An Objection

Before we embrace this conclusion, I want to consider the following objection: *Impact Constraint* and *Weighing Constraint* entail that only p is a reason to believe p . But that contradicts a number of independently plausible views about the nature of reasons. For example, according to the *Reasons-as-Evidence View*, reasons to ϕ are evidence that one ought to ϕ .³⁵ On the *Reasons-as-Explanations View*, reasons to ϕ are considerations that explain why one ought to ϕ . And on the *Reasons-as-Premises View*, reasons to ϕ are premises of good reasoning from these considerations toward ϕ -ing.³⁶ On all these views, independent evidence for p is a reason to believe p . If the two constraints imply otherwise, then so much the worse for these constraints.³⁷

Since it will be clear how my response generalizes to the other cases, I will make it explicit only for the case of the *Reasons-as-Evidence View*. The challenge is this: if the *Reasons-as-Evidence View* is correct, then independent evidence for p indicates that p is true, and so that one ought to believe p , and so, via the proposed analysis, that it is a reason to believe p . Since the two constraints deny this, they are inconsistent with an attractive view about the nature of reasons and so cannot support an argument against objectivism, at least without further defense.

The problem with this objection is that it begs the question against the proposed argument. There is a conflict between the two constraints and the independently plausible *Reasons-as-Evidence View* only if one already presupposes an objectivist reading of it. More precisely, the *Reasons-as-Evidence View* entails that both p and independent evidence for p are reasons to believe p only if we already

point is not that it is *conceptually* impossible to weigh up p and independent evidence bearing on p . One might think this to be the case, since weighing up p in doxastic deliberation could seem to imply that one already believes p , so that all cases where one weighs up p in deliberation are cases where one does not deliberate about whether to believe p (since, in these cases, one has already answered this deliberative question). But one can also weigh up considerations in deliberation that one only *supposes* to be true. In these cases, appealing to a consideration does not imply already believing it, and so it seems conceptually possible to weigh it up against other reasons. But there is still no deliberative context where this would be *coherent*.

35 E.g., Kearns and Star, "Reasons as Evidence."

36 E.g., Broome, "Reasons"; Way, "Reasons as Premises of Good Reasoning."

37 Thanks to Benjamin Kiesewetter and an anonymous referee for pressing me on this point.

agree that one must not impose any epistemic constraints on which facts can count as reasons for belief.

Consider, for example, a perspectivist version of this view, which says that for R to be a reason to believe p is for R to be evidence that one ought to believe p , of the kind that could rationally figure in the agent's deliberation given her evidence. Since, at the time of deliberating about whether to believe p , p could not rationally figure in such deliberation—at that point, whether p is precisely what one is about to find out—it follows from a perspectivist construal of the *Reasons-as-Evidence View* that only independent evidence is a reason to believe p . So understood, there is no conflict with the two constraints, which rule out only that both p and independent evidence for p are reasons to believe p . The conflict arises only if we start with an objectivist construal of the *Reasons-as-Evidence View*, which imposes no restriction whatsoever on which pieces of evidence can be reasons for belief. Yet in assessing these constraints, we cannot just presuppose an objectivist construal of the *Reasons-as-Evidence View*, since whether that view should be understood in an objectivist way or not is the very question currently at issue.

To put the same point differently: endorsing the two constraints does not contradict the *Reasons-as-Evidence View*, as long as one does not build an objectivist reading right into it. If one is sympathetic to the *Reasons-as-Evidence View*, the fact that an *objectivist version* of it would conflict with two independently plausible constraints about reasons is thus, if anything, a reason against endorsing an objectivist version of this view. The proposed conflict is illusory, unless we already assume that the only way for the *Reasons-as-Evidence View* to be true is for an objectivist version of it to be true, in which case we are begging the question against perspectivism. While I lack the space to go through this line of response for the other cases, it should be clear that the strategy generalizes. In all these cases, the proposed views conflict with the two constraints only if one already assumes that the right way to understand these views is along objectivist lines.

We should conclude that, given two independently plausible constraints about reasons, which are also compatible with a number of plausible views about the nature of reasons, objectivism implies that only p is a reason to believe p . This does not mean, of course, that objectivists must deny that independent evidence for p can play other normatively significant roles. For example, such evidence might explain why people are sometimes not blameworthy for their false beliefs. Still, objectivists have to deny that such evidence is ever a normative reason to believe something.

4.5. *How Objectivism Makes Correct Guidance Impossible*

The first premise of the Infection Argument says that epistemic objectivism rules out the possibility of believing normative propositions for the right reasons. As we have seen, if epistemic objectivism is true, then the only reason to believe p , and in fact the only *right* reason to believe p , is that p is true. Epistemic objectivism asserts this for all propositions. But, as I will now argue, the truth of a normative proposition (as opposed to independent evidence for it) is not a reason *for which* one could possibly believe that proposition.

There are several arguments for this claim. First of all, it is supported by widespread assumptions in the extant literature on the epistemic basing relation. For example, Ian Evans seems to assume that there is in fact *no* proposition, let alone a normative proposition, that could be believed on the basis of its own truth:

Very many of our beliefs are based on other beliefs. Some are based on perceptual experiences and some are based on apparent memories. Others are based on desires, fears, vanity, prejudices and other epistemically disreputable states. Perhaps some of our beliefs are baseless... A pervasive thought is that for a belief to be justified by some evidence, it is at least necessary that it be based on that evidence.³⁸

Evans here lists a number of candidate examples for the kinds of things that a belief could be based on. What we do not find among his suggestions is the truth of the proposition believed. Similarly, Allen Korcz's *Stanford Encyclopedia* entry on epistemic basing never explicitly mentions the possibility that p 's truth can be the reason for which someone believes that p ; the same goes for Kurt Sylvan's survey article.³⁹

Furthermore, there are influential theories of the epistemic basing relation that directly contradict this possibility. For example, Paul Moser's causal theory assumes the following necessary condition:⁴⁰

Moser's Condition: A belief that p is based on q only if the belief that p is (non-deviantly) causally sustained by the belief that q .

According to this condition, a belief that p would be based on p itself only if the belief that p causally sustained itself. But unlike God, perhaps, a belief is not a causally self-sustaining entity. It thus follows from the condition that a belief that p is not possibly based on p itself.

38 Evans, "The Problem of the Basing Relation," 2943.

39 Korcz, "The Epistemic Basing Relation"; Sylvan, "Epistemic Reasons II."

40 Moser, *Knowledge and Evidence*.

Similarly, Evans's dispositional theory explicitly assumes that "beliefs can only be based on other mental states."⁴¹ This immediately entails that one cannot believe *p* on the basis of *p*'s truth, since *p*'s truth is not a mental state. In general, it seems that only certain doxastic theories, according to which *S*'s belief that *p* is based on *q* if and only if *S* has an appropriate meta-belief that *q* is a good reason to believe *p*, are at least compatible with the possibility of believing *p* for the reason that *p*.⁴² But doxastic theories are controversial, and not all versions of such a view are consistent with this possibility.⁴³

The extant literature on epistemic basing hence does not leave much room for the claim that the epistemic objectivist's reasons to believe normative propositions are reasons for which people could believe such propositions. Since a detailed discussion of this literature is beyond the scope of this paper, I will not go into further details here. Instead, I will offer three independent reasons for thinking that it is impossible to believe a proposition for what is, according to objectivism, the only right reason to believe it.

First, consider considerations of theoretical simplicity. Everyone, even epistemic objectivists, must accept that at least some normative beliefs are based on independent evidence for these beliefs. For example, you could believe that *p* while having forgotten the reason for which you believe *p*. But then the reason for which you believe that *p* cannot be that *p*, since you cannot at the same time believe that *p* and have forgotten that *p*. Alternatively, consider false beliefs. Your false belief that killing other people for fun is morally permissible cannot be held for the reason that killing other people for fun is morally permissible, since killing other people for fun is not morally permissible. But if some truth-related normative beliefs are based on independent evidence, it is just more economical to assume that this is true for all such beliefs, rather than assuming that some normative beliefs are based on their truth whereas others are based on independent evidence for their truth.⁴⁴

Second, assuming that one could believe a normative proposition for the reason that it is true commits objectivists to a very strong form of intuitionism, according to which we can stand in a relation of direct acquaintance to norma-

41 Evans, "The Problem of the Basing Relation," 2945.

42 See, e.g., Tolliver, "Basing Beliefs on Reasons."

43 Some versions of the doxastic view deny that appropriate meta-beliefs are sufficient for a basing relation to obtain. Since these versions include additional causal conditions, they rule out that a belief that *p* could be based on *p* for the reason given above. Cf. Korcz, "The Epistemic Basing Relation," sec. 3.

44 By "truth-related" beliefs I mean beliefs that are not held for pragmatic reasons. I leave it open whether there are any beliefs that are not truth related in this sense.

tive facts. It should be clear that this is a very strong commitment, which invites all the problems that come with this sort of intuitionist normative epistemology. To be sure, such an epistemology might be quite plausible with respect to some classes of propositions, such as phenomenal propositions like "I have a pleasurable experience right now." Perhaps we can be directly acquainted with such truths; after all, *it seeming to me* that I have a pleasurable experience just is for me to *have* such an experience. So perhaps I cannot be mistaken about my having it, in which case it might seem plausible to say that I am directly acquainted with the fact of my having it or, in other words, that my believing to have this experience is based on my really having the experience.

But the case of normative truths is very different. In the normative case, it seeming to me that p does not make it the case that p , since my belief might be mistaken despite the seeming. So the reason for thinking that I am directly acquainted with p that applied in the phenomenal case does not apply here. This makes it all the more doubtful that normative propositions are among the propositions that could be believed on the basis of their truth. In general, it would clearly be better for a view about the perspective dependence of "ought" not to commit one to such controversial claims in normative epistemology.

Third, consider the psychology of false belief. To ask whether believers can acquire or maintain a belief on the basis of its truth is to ask whether truth is a reason *for which* one could believe. To ask whether something is a reason *for which* one could believe is to ask a question about psychology. Note, however, that the psychology of acquiring or maintaining false normative beliefs does not seem to differ *in kind* from the psychology of acquiring or maintaining true normative beliefs. My belief that p when p is true does not seem to differ psychologically, just because it is true, from my belief that p when p is false.

To illustrate, let us say you believe that killing other people for fun is morally impermissible. Now suppose that I, unlike you, got things horribly wrong about morality, believing that killing other people for fun is morally permissible. I might believe this because an ill-meaning authority told me so, or because I have been brainwashed. But it is certainly possible that no such thing occurred, and that I hold my belief because it seems as self-evidently true to me as your belief seems to you. In such a case, the psychological processes that led me to forming my false belief seem to be of the same kind as the psychological processes that led you to forming your true belief. Since the reason for which someone holds a belief is a matter of the psychological processes that led her to forming it, the reason for which I hold my belief thus seems to be of the same kind as the reason for which you hold your belief. Since my belief is actually false and so the reason for which I hold it cannot be that it is actually true, the same must then be

true for your belief. Instead, both beliefs are held for the reason that they seem self-evident to the person holding the belief. It is just that what is self-evident to one person need not be self-evident to another, and in this case, I was unlucky enough for an appearance of self-evidence to lead me astray.

In general, false normative beliefs are acquired and maintained for all sorts of reasons, including various types of evidence, superstition, habit, prejudice, or self-deception. There is just one reason for which a false normative belief cannot be held: its actually being true. But if a false normative proposition cannot be believed for the reason that it is true, and if the psychology of acquiring or maintaining true normative beliefs does not differ in kind from the psychology of acquiring or maintaining false normative beliefs, then it seems that a true normative proposition cannot be believed for the reason that it is true, either. True normative beliefs must instead be acquired or maintained for those other kinds of reasons, such as independent evidence.

What does this mean for the Infection Argument? We have seen that, if epistemic objectivism is true, only the truth of p is a reason to believe p . But as the above arguments show, p 's truth—for some normative proposition p —is not a reason for which one could believe p . Hence, epistemic objectivism rules out that people ever believe normative propositions for the reason which, according to objectivism, is the right reason to believe such a proposition, viz., that it is true.

Now recall condition B2 in my earlier analysis of correct action guidance. For correct action guidance to be possible, it must be possible for an agent to perform the action required by a theory on the basis of correct practical reasoning from that theory. This, in turn, requires not just that the agent believes, for some R that decisively favors the action according to the theory, that R decisively favors the theory, but also that she believes *this* for the right reason.

However, that is precisely what is ruled out by objectivism. To return to our earlier example, if objectivism holds in both the practical and the epistemic domain, then the right reason to believe that promises generate decisive reasons according to Kantianism—now assuming an objectivist version of Kantianism—is that the categorical imperative *actually* entails this. In other words, it is the truth of the proposition “the categorical imperative entails such-and-such” that is the right reason to believe this proposition and not, as in a perspectivist version of Kantianism, that the *available evidence* suggests that the categorical imperative entails this.

As we have seen, however, the fact that the categorical imperative actually entails X is not a reason for which one could believe that it entails X . If an objectivist version of Kantianism is true, the only right reason to believe that promises

generate decisive reasons is hence a reason for which it is impossible to believe that promises generate decisive reasons. If epistemic objectivism is true, it is thus impossible for B2 to be satisfied, and so impossible to act on the basis of correct practical reasoning from Kantianism toward the action required by Kantianism. Ultimately, if objectivism is true, it is impossible to be correctly guided by that theory (*mutatis mutandis* for normative theories other than Kantianism).

Perspectivism, on the other hand, avoids this result: that there is sufficient evidence for the proposition “the categorical imperative entails such-and-such” is a right reason to believe that proposition according to perspectivism, and it is also a reason for which one could believe that proposition.

I conclude that, if objectivism is true, then it is impossible to believe a normative proposition for the right reason. This establishes premise 1 of the Infection Argument, which in turn establishes the crucial premise of the Argument from Action Guidance: that, if objectivism is true, it is impossible for someone’s action to be correctly guided by a normative theory. Since that is implausible, so is objectivism.

5. CONCLUSION

In this essay, I have provided a new guidance-based argument against objectivism about the practical “ought.” This argument draws together two separate themes from the recent literature: first, that the normative is in some sense action guiding, and second, that the parallel between practical and epistemic normativity provides fruitful resources for debating the plausibility of objectivism and competing views.

The main idea behind the argument was to show that if objectivism is true then, implausibly, it is impossible for an action to be correctly guided by a normative theory. In defense of this argument, I developed a substantive account of what it is for a normative theory to correctly guide action. The crucial condition of this analysis says that correct guidance requires the capacity to believe normative propositions for the right reasons. But since practical objectivism entails epistemic objectivism, and epistemic objectivism rules out this capacity, it follows that correct guidance is impossible.

Not only is this an unacceptable result, it is also a result that perspectivism easily avoids. This provides a robust, theory-driven argument against objectivism, which does not just appeal to contested intuitions about particular cases but

also sheds light on a notion of action guidance that should be of independent interest to normative theory.⁴⁵

Humboldt University of Berlin
philip.fox@hu-berlin.de

REFERENCES

- Andrić, Vuko. "The Case of the Miners." *Journal of Ethics and Social Philosophy* 7, no. 1 (January 2013): 1–8.
- . "Objective Consequentialism and the Licensing Dilemma." *Philosophical Studies* 162, no. 3 (February 2013): 547–66.
- . "Objective Consequentialism and the Rationales of 'Ought' Implies 'Can.'" *Ratio* 30, no. 1 (March 2017): 72–87.
- Björnsson, Gunnar, and Stephen Finlay. "Metaethical Contextualism Defended." *Ethics* 121, no. 1 (October 2010): 7–36.
- Boghossian, Paul. "Is Meaning Normative?" In *Content and Justification: Philosophical Papers*, 95–108. Oxford: Oxford University Press, 2008.
- Broome, John. *Rationality through Reasoning*. Oxford: Wiley Blackwell, 2013.
- . "Reasons." In *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, edited by Jay Wallace, Michael Smith, Samuel Scheffler, and Philip Pettit, 28–55. Oxford: Oxford University Press, 2004.
- Bykvist, Krister. "How to Do Wrong Knowingly and Get Away with It." In *Neither/Nor: Philosophical Papers Dedicated to Erik Carlson on the Occasion of His Fiftieth Birthday*, edited by Rysiek Sliwinski and Frans Svensson, 31–47. Uppsala, Sweden: Uppsala Philosophical Studies, 2011.
- Evans, Ian. "The Problem of the Basing Relation." *Synthese* 190, no. 14 (September 2013): 2943–57.
- Gibbard, Allan. "Truth and Correct Belief." *Philosophical Issues* 15 (2005): 338–50.
- Gibbons, John. *The Norm of Belief*. Oxford: Oxford University Press, 2013.

45 For their helpful comments on earlier versions of this article, I would like to thank Daniele Bruno, Jonas Burgstahler, Jan Grohn, Marcel Jahn, Benjamin Kiesewetter, Thomas Schmidt, Jonathan Way, the participants of Thomas Schmidt's ethics colloquium, and the editor of this journal, Mark Schroeder, along with two anonymous referees. I would also like to thank the participants of a graduate seminar on ignorance and moral responsibility by Thomas Schmidt and Daniele Bruno at Humboldt University, Berlin, for their comments on some of the previous material that made it into this article. Work on this project has been supported by the Studienstiftung des deutschen Volkes.

- Graham, Peter A. "In Defense of Objectivism about Moral Obligation." *Ethics* 121, no. 1 (October 2010): 88–115.
- Henning, Tim. "Normative Reasons Contextualism." *Philosophy and Phenomenological Research* 88, no. 3 (May 2014): 593–624.
- Hills, Alison. "What Does It Take to Act for Moral Reasons?" In *The Many Moral Rationalisms*, edited by Karen Jones and François Schroeter, 247–63. Oxford: Oxford University Press, 2018.
- Jackson, Frank. "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101, no. 3 (April 1991): 461–82.
- Kearns, Stephen, and Daniel Star. "Reasons as Evidence." *Oxford Studies in Metaethics*, vol. 4, edited by Russ Shafer-Landau, 215–42. Oxford: Oxford University Press, 2009.
- Kiesewetter, Benjamin. "How Reasons Are Sensitive to Available Evidence." In *Normativity: Epistemic and Practical*, edited by Conor McHugh, Jonathan Way, and Daniel Whiting, 90–114. Oxford: Oxford University Press, 2018.
- . *The Normativity of Rationality*. Oxford University Press, Oxford, 2018.
- . "'Ought' and the Perspective of the Agent." *Journal of Ethics and Social Philosophy* 5, no. 3 (October 2011): 1–24.
- . "What Kind of Perspectivism?" *Journal of Moral Philosophy* 15, no. 4 (2018): 415–43.
- Kolodny, Niko. "How Does Coherence Matter?" *Proceedings of the Aristotelian Society* 107, no. 3 (October 2007): 229–63.
- Kolodny, Niko, and John MacFarlane. "Ifs and Oughts." *Journal of Philosophy* 107, no. 3 (March 2010): 115–43.
- Korcz, Keith Allen. "The Epistemic Basing Relation." *Stanford Encyclopedia of Philosophy* (Fall 2015). <https://plato.stanford.edu/archives/fall2015/entries/basing-epistemic>.
- Littlejohn, Clayton. "Do Reasons and Evidence Share the Same Residence?" Unpublished manuscript.
- . *Justification and the Truth-Connection*. Cambridge: Cambridge University Press, 2012.
- . "Prichard's Question and Reasons Perspectivism." Unpublished manuscript.
- . "The Unity of Reason." In *Epistemic Norms: New Essays on Action, Belief, and Assertion*, edited by Clayton Littlejohn and John Turri, 135–54. Oxford: Oxford University Press, 2014.
- Lord, Errol. "Acting for the Right Reasons, Abilities, and Obligation." In *Oxford Studies in Metaethics*, vol. 10, edited by Russ Shafer-Landau, 26–52. Oxford: Oxford University Press, 2015.

- . “The Real Symmetry Problem(s) for Wide-Scope Accounts of Rationality.” *Philosophical Studies* 170, no. 3 (September 2014): 443–64.
- McHugh, Conor, and Jonathan Way. “Objectivism and Perspectivism about the Epistemic Ought.” *Ergo* 4, no. 5 (2017): 121–45.
- Moore, G. E. *Ethics*. New York: H. Holt and Co., 1912.
- Moser, Paul K. *Knowledge and Evidence*. Cambridge University Press, Cambridge, 1989.
- Piller, Christian. “Ewing’s Problem.” *European Journal of Analytic Philosophy* 3, no. 1 (2007): 43–65.
- Pittard, John, and Alex Worsnip. “Metanormative Contextualism and Normative Uncertainty.” *Mind* 126, no. 501 (January 2017): 155–93.
- Prichard, H. A. “Duty and Ignorance of Fact.” In *Moral Writings*, 84–101. Oxford: Clarendon Press, 2002.
- Robertson, Simon. “Epistemic Constraints on Practical Normativity.” *Synthese* 181, supplement 1 (August 2011): 81–106.
- Ross, W. D. *The Right and the Good*. Oxford: Clarendon Press, 1930.
- Scanlon, Thomas. *What We Owe to Each Other*. Cambridge, MA: Belknap Press of Harvard University Press, 1998.
- Schroeder, Mark. “Knowledge Is Belief for Sufficient (Objective and Subjective) Reasons.” In *Oxford Studies in Epistemology*, vol. 5, edited by Tamar S. Gendler and John Hawthorne, 226–52. Oxford: Oxford University Press, 2015.
- . “Getting Perspective on Objective Reasons.” *Ethics* 128, no. 2 (January 2018): 289–319.
- Sepielli, Andrew. “Subjective Normativity and Action-Guidance.” *Oxford Studies in Normative Ethics*, vol. 2, edited by Mark Timmons, 45–73. Oxford: Oxford University Press, 2012.
- Shah, Nishi. “How Truth Governs Belief.” *Philosophical Review* 112, no. 4 (October 2003): 447–82.
- Shah, Nishi, and J. David Velleman. “Doxastic Deliberation.” *Philosophical Review* 114, no. 4 (October 2005): 497–534.
- Sylvan, Kurt. “Epistemic Reasons II: Basing.” *Philosophy Compass* 11, no. 7 (July 2016): 377–89.
- Thomson, Judith Jarvis. “Imposing Risks.” In *Rights, Restitution, and Risk*, 173–91. Cambridge, MA: Harvard University Press, 1986.
- . *Normativity*. Chicago: Open Court, 2008.
- Tolliver, Joseph. “Basing Beliefs on Reasons.” *Grazer Philosophische Studien* 15, no. 1 (1982): 149–61.
- Way, Jonathan. “Reasons as Premises of Good Reasoning.” *Pacific Philosophical Quarterly* 98, no. 2 (June 2017): 251–70.

- . “The Symmetry of Rational Requirements.” *Philosophical Studies* 155, no. 2 (September 2011): 227–39.
- Way, Jonathan, and Daniel Whiting. “Perspectivism and the Argument from Guidance.” *Ethical Theory and Moral Practice* 20, no. 2 (April 2017): 361–74.
- Wedgwood, Ralph. “The Aim of Belief.” *Philosophical Perspectives* 16 (2002): 267–97.
- Weiner, Matthew. “The Spectra of Epistemic Norms.” In *Epistemic Norms: New Essays on Action, Belief, and Assertion*, edited by Clayton Littlejohn and John Turri, 201–18. Oxford: Oxford University Press, 2014.
- Whiting, Daniel. “Should I Believe the Truth?” *Dialectica* 64, no. 2 (June 2010): 213–24.
- Zimmermann, Michael J. *Ignorance and Moral Obligation*. Oxford: Oxford University Press, 2014.
- . *Living with Uncertainty: The Moral Significance of Ignorance*, Cambridge: Cambridge University Press, 2008.

BASICALLY DESERVED BLAME AND ITS VALUE

Michael McKenna

HOW SHOULD WE understand basic desert as a justification for blaming? Many philosophers account for free will by identifying it with the control condition for *basic desert-entailing moral responsibility*.¹ On such a view, a blameworthy person deserves blame just because of how she acted—for instance, because she knowingly and freely acted morally wrong. Crucially, the justification provided by desert is *not* rooted in any other normative consideration, such as utility, or the reasonableness of entering into a contract with others.² But what precisely does basic desert come to? And what is it about blame that makes it the thing that a blameworthy person deserves? Moreover, how is any particular instance of blame fitted properly—rather than ill fitted—for a blameworthy person's particular act so that it is the thing that is deserved? As it turns out, there are challenges to understanding basic desert for blame, challenges having nothing to do with skepticism about free will. One challenge concerns whether the only good in harming a person by blaming her is exclusively instrumental. Another challenge concerns traditional worries about retributivist theories of punishment that might threaten deserved blame too. Given these challenges, there may be reason to reject desert-based conceptions of the justification of blame for reasons altogether distinct from any worries about free will.

In what follows, by drawing upon my own conversational theory of moral responsibility, I will develop the view that blame is to be justified in terms of basic desert.³ I have three interrelated aims. One is to account for the fittingness of blame on analogy with the fittingness of a move in an actual conversation between competent linguistic practitioners of the same language. Another is to solve the problem of explicating the desert relation regarding what is deserved by the blameworthy in a way that helps avoid traditional worries about retributive

1 E.g., Fischer and Ravizza, *Responsibility and Control*; Mele, *Free Will and Luck*; Pereboom, *Free Will, Agency, and Meaning in Life*; and Sartorio, *Causation and Free Will*.

2 E.g., Feinberg, *Doing and Deserving*; Pereboom, *Living Without Free Will and Free Will, Agency, and Meaning in Life*; and Scanlon, *Moral Dimensions*.

3 McKenna, *Conversation and Responsibility*.

theories of punishment. The third is to defend the controversial thesis that the harm involved in blaming can be good in a way that is not merely instrumental.

In developing this view, I mean to articulate an account of basic desert-entailing moral responsibility that is neutral between freewill realists and freewill skeptics. This should prove useful in helping adjudicate the debate between those realists and skeptics who agree that what is in dispute between them is the freedom required to deserve blame in a basic sense. I will restrict myself just to blameworthiness and blame, leaving aside praiseworthiness and praise, as well as moral responsibility for conduct that is neither praiseworthy nor blameworthy.

I begin with two preliminary qualifications. First, the instances of blame I take to be of interest, and the role basic desert plays in justifying them, have to do with *directed blame*, wherein those blaming direct their blame toward the blamed person by overt means.⁴ In such cases, the blamed person is positioned to recognize the blame and register that those blaming her intend that she receive their blame. Second, I assume that, at least in paradigmatic cases, directed blame harms the person blamed. For this reason, I will assume that one cannot justify the goodness in blaming a person unless one can also justify the harms that attend blame.⁵

1. BASIC DESERT

With the preceding clarifications in place, consider basic desert. What is it? Given what has already been stated, we know that it is basic at least in the following negative way: the normative warrant it provides is *not* supported by any more fundamental normative principles or values.⁶ But that is not saying much. What more can be said to give some positive content to our understanding of it? I will restrict my attention to deserved moral blame. I will not concern myself with

4 McKenna, "Directed Blame and Conversation."

5 I take it as a fact about the sort of blame at issue that it is liable to harm in various ways, and that the normative burden of offering a justification for blaming requires a justification for these sorts of harms. Of course, there are antiseptic notions of blame, like Smart's notion of dispraise, which are meant to be washed clean of any harmful effects ("Free-Will, Praise and Blame"). But most assume, as Pamela Hieronymi puts it, that blame directed at a blamed person has a force or a characteristic sting, something painful or burdensome ("The Force and Fairness of Blame"). Part of the philosophical project of offering a theory of moral responsibility and of justified blame requires accounting for and justifying this. I am grateful to a referee from this journal for asking me to clarify this point.

6 E.g., see Feinberg, *Doing and Deserving*, 56; Pereboom, *Free Will, Agency, and Meaning in Life*, 2; and Scanlon, *Moral Dimensions*, 188.

basic desert in other domains, like the desert for winning a prize or being treated with respect as a person.

To begin, desert offers a distinctive way of specifying the *sense of aptness* in a judgment that blame is appropriate.⁷ Mere aptness on its own simply reports *that* some normative warrant exists; it gives no content to the kind of warrant on offer. Desert does so, not by appealing to considerations of utility, or principles of fairness, or the elements of a reasonable contract, but by appealing exclusively to a “desert-base” that makes fitting that which is deserved.⁸ So understood, desert is a distinctive species of *fittingness*. As regards deserved blame, the desert-base for a blameworthy act involves only salient features of *the agent* and *her act*, features that make the agent blameworthy for it. Suppose, to make things simple, this consists just in an agent knowingly and freely doing morally wrong. Here we have three ingredients contributing to the desert-base: one concerns the agent’s state of knowledge, another concerns the agent’s relation to the act as a free one, and the third concerns the moral status of the act itself as one that is morally impermissible. These features provide the desert-base for a response that is fitted for the agent’s act—in particular, a blaming response.

The blaming response is meant to fit the act in relation to the features of the desert base in some unique, case-specific manner, one that is especially difficult to specify. (I offer a proposal below.) There is, furthermore, as noted above, a widely shared presumption that blaming is negative in a way that involves exposing the one blamed to the liability of certain harms; it has a characteristic sting or force. Crucially, the blame’s being basically deserved *exhausts* the requisite positive normative warrant for exposing the blameworthy agent to such harms.

Why write in terms of basic desert exhausting the *positive* normative warrant for blaming? On a credible version of a basic-desert thesis, the complete normative warrant for actually blaming an agent, one yielding an all-out judgment, also requires the *negative* condition that there are no competing and overriding normative considerations, like those of overall utility or simple prudence, speaking against blaming. Hence, basic desert only provides *pro tanto* reasons. So it does not immediately follow that if it is true that a person deserves blame, in the all-out sense the right thing for someone (or other) to do is blame that person.

To help give further content to judgments of deserved blame, it is useful to consider whether they involve only the right or instead also the good.⁹ There is

7 Feinberg, *Doing and Deserving*, 56–57.

8 Feinberg, *Doing and Deserving*, 58–61.

9 In what follows, I will seek an elucidation of basic desert by considering whether a claim of basic desert entails a judgment that is either deontic or axiological. One referee for this journal expresses skepticism for this strategy. About the proposal I will advance in this para-

no consensus on this point. Some have in mind an exclusively deontic rendering. For example, as Joel Feinberg put it:

That a subject deserves *X* entails that he ought to get *X* in the *pro tanto* sense of “ought.”¹⁰

On such a view, there is no entailment from its being deserved to its being *good* that a blameworthy wrongdoer is blamed and thereby harmed. Indeed, in developing his own “*desert-based view*,” Scanlon explicitly denies this:

The fact that someone has behaved wrongly can make it appropriate to withhold certain attitudes and relationships, and withholding these things may make the person’s life worse. But withholding them is justified, in my view, by the fact that they have become inappropriate, not by the fact that withholding them makes the person worse off. Ceasing to hope that things go well for a person can be one element of blame, but as I have emphasized, this does not involve thinking it to be good that things not go well for him.¹¹

On a strong version of an exclusively deontic desert thesis, blaming one who deserves blame would be construed as a duty or an obligation. On a weak rendering, it would be cast simply as something that is permissible.¹²

Others favor an axiological thesis that supplies the basis for a deontic judgment. On such a view, the goodness of the harm in blaming provides a justification for, at a minimum, the permissibility of doing so.¹³ In advancing such a view, Christopher Bennett writes:

graph, they write, “Put this way, the view does not appear to answer a question that [the author] raises at the start: what is basic desert? An entailment of basic desert is identified, but identifying an entailment of it is not saying what it is. Basic desert seems to be left as a something-or-other that entails this value.” In response, I do not think this is quite right. While it is true that, in some cases, we do not learn what *x* is by learning what is entailed by *x*. But sometimes we do. We learn something about a dolphin when we learn that it is a mammal, and so it is informative to learn that “Flipper is a dolphin” entails “Flipper is a mammal.” The salient feature of my philosophical strategy is to seek some non-reductive means of elucidating desert when it is regarded as basic and so not grounded in any more basic normative notion. I am treating the entailments identified here as markers of essential features of desert wherein desert is the ground for either the rightness or goodness of the deserved thing. (I wish to emphasize here my gratitude to this referee for pressing me on this and several other matters in the following couple of paragraphs.)

10 Feinberg, *Doing and Deserving*, 60.

11 Scanlon, *Moral Dimensions*, 188.

12 E.g., Scanlon, *Moral Dimensions*, 188–89.

13 E.g., Bennett, “The Varieties of Retributive Experience”; and McKenna, “Directed Blame

I shall show the extent to which our participation in the reactive attitudes [ones expressive of blaming] betrays a commitment to retribution, to the thought that it is a non-contingently good thing that those who have done wrong should undergo some form of suffering.¹⁴

Without defending the view, R. Jay Wallace characterizes the thesis of retributivism similarly as

the view that it is intrinsically good that wrongdoers should suffer harm, and that therefore we have a positive duty to inflict such harms on them.¹⁵

Here, I take Wallace's formulation of retributivism, like Bennett's, to be an instance of an axiological basic-desert thesis.

In what follows, I will advance a variation on an axiological thesis. A weak version would treat the goodness of blaming as a justification for the mere permissibility of blaming, whereas a strong version would have it as a moral requirement, as for example it is expressed in the preceding quotation from Wallace.¹⁶ A middle ground, which I now endorse, is that the goodness of the harm in blaming provides a practical reason *favoring* blaming.¹⁷ Favoring practical reasons are of an intermediate strength as between reasons issuing from requirements and the merely weak reason of permissibility. As I understand them, all that is provided by considerations of permissibility is simply *that* nothing prohibits a certain course of action. Favoring reasons for blaming seem best suited for a thesis about desert. How so? Contending that those positioned to blame have a moral duty to blame is overly demanding to account for our sense of the option of blame in a wide range of cases. But mere permissibility does not capture the force of our reasons *to* blame. Favoring does. It is plausible, granting the axiological assumption, that if it is in some way good to harm a person by blaming her, the goodness counts as a reason favoring such a course of action, while at the same time not requiring it.

But why adopt *any* axiological desert-based account of blame? Why not commit to an exclusively deontic version? Because the latter commits to less—in particular, because it makes no commitment to the goodness of the harm in blaming—is it not philosophically easier to defend? Here is an argument. A strong version of an exclusively deontic thesis is a nonstarter for blame. Re-

and Conversation," 133.

14 Bennett, "The Varieties of Retributive Experience," 147.

15 Wallace, *Responsibility and the Moral Sentiments*, 60n13.

16 E.g., Bennett, "The Varieties of Retributive Experience"; McKenna, "Directed Blame and Conversation"; Wallace, *Responsibility and the Moral Sentiments*, 60n13.

17 I am indebted to Ingmar Person for this suggestion.

quiring others to blame directly the blameworthy places unreasonably high demands on members of the moral community, and most notably those who have been wronged. But a weak version, merely rendering it permissible to blame the blameworthy, only establishes that one would do no wrong to blame. This provides no positive reason *to* blame. It offers nothing that favors blaming as one of the constitutive features of a desert-based intuition.¹⁸ Insofar as directed blame involves harm, this sort of normative warrant for directed blaming remains silent on why anyone *should* do this harm-causing thing. But if blaming involves harming, it seems one should *not* blame unless she has a good reason to do so; the harming would itself seem to offer a reason *against* blaming, even if it is permissible to harm. So a culpable person's deserving blame would never outweigh a would-be blamer's reasons to refrain from blaming.¹⁹ That cannot be right. An axiological version that favors blaming because it is good supplies the sort of reason at issue and so avoids this problem.²⁰

One might protest that the preceding argument relies upon a mistaken assumption about what normative work an appeal to basic desert should do. So, the objection might go, basic desert only needs to justify a wrongdoer receiving blame. It need not provide the good reasons one blaming has *to* blame the per-

- 18 A referee for this journal asks why the reason for blaming is not provided merely by the desert base, absent a further presumption that blaming is good. Why, this referee asks, is it not enough of a reason to blame a person that she freely and knowingly acted wrongly, for instance? In a sense, it is. The thesis under consideration is what a claim of desert comes to, of what desert is (as it bears on blame). The claim here, on the imagined weak deontic thesis, is that the desert base provides no more than the following as a reason: that it is permissible to blame. This is tantamount to the weak thesis that, *pro tanto*, it would not be wrong to blame. My argument is that this is inadequate, since the *harm* in blaming would then supply a reason (not a requiring reason) not to blame.
- 19 I am grateful to an anonymous referee for suggesting this way of formulating the point.
- 20 A referee for this journal has asked whether a deontic thesis cast in terms of *pro tanto* duties would do work similar to the work I contend is best done by favoring reasons on an axiological approach. Would these weaker duties not give rise to a weak deontic thesis, one that would not be any different from my proposed axiological proposal? I do not think so. *Pro tanto* duties are still duties, and duties are requirements that, when one fails them, involve wrongdoing. Of course, being *pro tanto*, they can admittedly be overridden. But absent overriding factors, they are still binding as duties. They are thus not optional for the blamer in a way that is best suited to most instances of warranted blame. Perhaps the preceding is too quick, as Mark Timmons has recently noted in conversation. There are further avenues a deontological theorist might consider. For instance, following a suggestion by David McNaughton (*Moral Vision*), one might distinguish moral oughts that are required from moral oughts that are simply recommendations. I grant that this is an alternative that one might pursue. If viable, the argument I have offered here would fall short. But this alternative proposal would need to be developed. Regardless, I will not pursue the matter here.

son who deserves it. Those reasons, it might be argued, are supplied by other factors, such as a wronged person's need to give voice to her own feelings of anger, or to aim at deterrence, or whatever. Perhaps one could regard these as favoring reasons. As such, one might argue, a strictly deontic version merely rendering it permissible to blame *could* make room for favoring reasons, and so, supplemented by these reasons, provide the practical reasons a blamer might have to do the blaming. But on such a view, strictly speaking, the desert itself, being an exclusively deontic notion, would *not* issue or be the ground of those favoring reasons.

I find the preceding defense of a deontic version an unsatisfying way to understand the role judgments about basically deserved blame play in our moral responsibility practices. In particular, the reasons favoring blaming when someone deserves blame in a basic sense are dependent on factors that go beyond those bearing on the desert itself. However, I will not pursue the matter any further here. Perhaps after all an exclusively deontic desert thesis *is* the philosophically wiser option insofar as it does commit to less. It must be granted that it would avoid the seemingly paradoxical issue to be discussed below: that harming itself could be good in a way that is not merely instrumental. Nevertheless, in what follows I will develop a favoring version of an axiological thesis. I offer two motives for pursuing this option, motives that remain even if the argument offered in the two preceding paragraphs proves unsuccessful.

First, there are some whose folk intuitions about desert include a consideration about the goodness and not just the rightness of a blameworthy person getting what she deserves. It is worth examining whether such a view can withstand careful scrutiny even if doing so is not strictly required to make sense of a basic-desert thesis for blame. Second, consider those philosophers who theorize about free will in terms of the control required for basic-desert-entailing moral responsibility. What I suspect many of them have in mind is a thesis about the *value* of agents getting what they deserve when, acting of their own free will, they knowingly do morally wrong. I am interested in engaging these philosophers and in developing for them a clear specification of what basically deserved blame comes to.²¹

2. ZEROING IN ON A PRINCIPLE OF BASIC DESERT

To illustrate the sort of view I will explore hereafter, here is a first pass at a principle restricted to morally wrong acts:

21 One of the referees from this journal has understandably asked for some support of this claim. I confess, I have none, save for some sense from talking with those working in this field.

D1: Because it is good to harm by blaming one who is blameworthy for a morally wrong act, there is a reason that favors doing so.

D1 does not identify or give the meaning of a desert thesis for blame. This is because it does not in any way express the fitting relation that is supposed to hold, by virtue of desert, between the desert-base involving an agent's morally wrong act and the blaming that is itself that which is deserved. D1 only reports the kind of appraisal—deontological or axiological—a claim of desert is. So, it is best to think of D1 as *entailed* by a certain desert thesis rather than identifying it or giving its (complete) meaning.²²

Note also that D1 is consistent both with a nondesert thesis, and also with a desert thesis that is *not* basic. To be clear, there are two distinct theoretical alternatives to a basic desert thesis one might consider when seeking to provide the normative warrant for blame. One is to offer an account specifying the appropriateness of blame that makes no reference to desert at all.²³ A different approach involves arguing that there are more basic normative principles providing the underlying rationale for claims of *what is deserved*, and hence for propositions such as D1. This would amount to a *nonbasic* desert thesis. How so? The good mentioned in D1 could be construed as an instrumental good, and then justified by reference to, for example, utilitarian or instead contractualist principles. On this approach, there would be important content to claims of desert that relate desert-bases to what is deserved in the proper sort of fitting manner. But the practices and underlying rationale for the claims of desert, and for a principle like D1, would themselves be grounded on something that is more fundamental.²⁴

Building upon D1, here is a revised principle more carefully suited just for basically deserved blame:

D2: Because it is intrinsically good to harm by blaming one who is blameworthy for a morally wrong act, there is a reason that favors doing so.

22 One way to construe the thesis, then, is that the goodness identified in D1 is *grounded* in desert. Careful attention from a referee for this journal helped me see the benefit in explaining the thesis in these terms.

23 Consider, for instance, a view like Smart, "Free-Will, Praise and Blame."

24 This is roughly what Rawls has called post-institutional desert (*A Theory of Justice*, 103–4). James Lenman, for instance, has explored an interesting nonbasic desert thesis for blame by appeal to contractualist considerations ("Compatibilism and Contractualism"). In my estimation, this is how one should understand Wallace, who rests claims of desert on considerations of fairness (*Responsibility and the Moral Sentiments*, 227). More recently, Manuel Vargas, *Building Better Beings*, has advanced a nuanced consequentialist alternative.

Some will object that harm to a person is never intrinsically good.²⁵ But to appreciate what those wishing to endorse D2 might have in mind, we can understand claims of intrinsic goodness in terms of cross-world differences. To illustrate, consider just for a moment not blame but instead punishment. Grant that it harms a person to sit in prison for five years; he is made to suffer in his isolation. Is this intrinsically bad? Suppose it is. But now consider instead a person who sits in prison for five years *in response to* having brutally raped someone. Here is a way to appreciate the claim of intrinsic goodness in D2: all things being equal, it is a better world that this rapist sits in prison for five years for his act of rape than a world in which this same rapist causes the same harm to his victim but undergoes no harm himself. That is fitting. Furthermore, in keeping with D2, because it is intrinsically good as a fitting response to the harm he caused, there is a reason that favors his being harmed.²⁶

4. MORAL RESPONSIBILITY, BLAMEWORTHINESS, AND BLAME

Given the preceding discussion of basic desert for blame, consider now blameworthiness and blame. Basic desert is alleged to provide the (or at least a) normative warrant for blame as a response to the blameworthy. To understand what basically desert-entailing moral responsibility is, we need an account of the pertinent *relata*: blameworthiness and blame. Here I summarize my own Strawsonian-inspired, communication-based theory.²⁷

To begin, I endorse the widely shared Strawsonian thesis that attention to a special class of emotions is integral to our understanding of moral responsibility.²⁸ Here our interest can be limited to the reactive attitudes of resentment

25 E.g., see Scanlon, "Giving Desert Its Due," 102.

26 Does this mean that it is not intrinsically bad that this criminal sit in prison for five years? No. It can be granted that incarceration is always an intrinsic bad. Still, that very thing, as a response to a crime, might be good as well along some other dimension. The appeal to a cross-world comparison between a world where the criminal is not punished and a world where he is helps to bring this out. I am indebted to a referee for this journal for encouraging me to clarify this point.

27 McKenna, *Conversation and Responsibility*. Several other philosophers have also endorsed and developed some variation on a communicative theory of moral responsibility, and much of what follows could be fitted for their views as well (e.g., Bennett, "The Varieties of Retributive Experience"; Macnamara, "Reactive Attitudes as Communicative Entities"; Shoemaker, *Responsibility from the Margins*; and Watson, "Responsibility and the Limits of Evil."

28 E.g., Strawson, "Freedom and Resentment"; Watson, "Responsibility and the Limits of Evil"; Wallace, *Responsibility and the Moral Sentiments*; Fischer and Ravizza, *Responsibility and Control*; Shoemaker, *Responsibility from the Margins*.

and moral indignation as they bear on blame. I note four features bearing on the discussion to follow.

First, an episode of one of these reactive attitudes is an affect-laden negative reaction *in response to* the presumption of another's objectionable quality of will, typically, but not always, for a wrong act. The response is in some way dependent upon, or normatively assessed in light of, a cognitive judgment of some sort, such as the judgment that an agent has done wrong.

Second, quality of will concerns the regard or lack of regard for others' interests and for salient moral considerations. Quality of will is not limited to the rightness or wrongness of an agent's act; it also includes, for example, her concern for or attitude regarding the rightness or wrongness of her act.

Third, the exemplar cases of these emotions, the ones that should be central to a theory of moral responsibility, are the ones that are publically manifested in outward behavior, and are directed at one who is blamed. These are the ones that pertain to what early on I identified as directed blame. Of course, there are private episodes of these emotions, but it is the overt ones, displayed in behavior, that are especially salient. More particularly, of these overt episodes, it is the directed ones, exhibited in the presence of the blamed and suitably directed as modes of blaming, that are crucial.²⁹

Fourth, the overt manifestations of these emotions are best understood as practice dependent. They involve alterations to interpersonal practices constitutive of adult interpersonal life. Such alterations to these practices help make manifestations of these emotions intelligible as the sorts of emotions they are.

Reacting to a blameworthy agent with a reactive attitude like resentment or moral indignation has an expressive and communicative role. This is a point Gary Watson emphasized in his development of a Strawsonian theory of moral responsibility.³⁰ In typical cases, an episode of resentment *expresses* an agent's moral concerns—her demands, expectations, regard for a blameworthy agent's moral violation, and, most importantly, her concern for the blameworthy agent's morally objectionable quality of will. Not only do manifestations of these emotions in our blaming practices express these things, they also serve a communicative role: they *communicate* our concerns by way of our altered regard toward the agent and her conduct in reaction to her blameworthy conduct. The manner of expression and communication is by way of altered interpersonal practices characteristic of the otherwise normal modes of conduct we engage in when dealing with each other. Normal courtesies might be withheld, friendships dis-

29 McKenna, "Directed Blame and Conversation."

30 Watson, "Responsibility and the Limits of Evil."

solved or attenuated, direct expressions of anger put on full display. All of these, in both word and deed, serve expressive and communicative roles.

Drawing upon the expressive and communicative nature of the morally reactive attitudes, I have proposed a bolder thesis: that the distinctive expressive and communicative nature of these attitudes—in their directed manifestations and within the context of a set of social practices—can take on something analogous to a conversational role. In particular, they can be understood on analogy with a stage in a conversation in which one responds to an interlocutor who has antecedently initiated a conversation. The basic proposal works like this: the actions of morally responsible agents are candidate bearers of a species of meaning—agent meaning—where meaning is a function of the quality of the will with which an agent acts. An agent's engaging in (seemingly) blameworthy conduct can be understood on analogy with a speaker of a natural language initiating a conversational exchange by way of saying something meaningful. When others then react by holding an agent morally responsible, by manifesting their reactive attitudes in cases of directed blaming, their responses can be understood as meaningful replies to the agent who initiated the "conversational exchange." When the (putatively) blameworthy agent then offers an apology, an excuse, a justification or admission of guilt, a defiant dismissal, or whatnot, this can be seen as a reply and thus a further extension of the evolving conversation.

One of the norms governing (the analog to) a conversational exchange will be informed by the meaning of the agent's action as indicative of the quality of will with which she acts. Some blaming responses will be especially suited as intelligible, even insightful, replies to the agent, just as sometimes, in conversation, one knows just the right thing to say in reply to her interlocutor so as to speak directly to exactly what was originally meant. Other blaming responses will be infelicitous insofar as they fail to appreciate the actual meaning of the agent's act. Indeed, it seems that the way excuses and justifications work attests to this point. In different ways, both show that an agent's action did not manifest the lack of moral regard for others that the content of a (sometimes preliminary) blaming response is premised upon. Also, the meaningfulness of a directed episode of resentment or instead moral indignation will be a function of the background social practices and expectations against which alterations will be seen to have a particular salience.

Here is an example I have used elsewhere.³¹ Two friends meet for coffee, and one, Daphne, is offended by a racist remark (she believes) Leslie makes about Hispanics. Daphne's indignation is then manifested in part by her not inviting Leslie to lunch as she normally does, along with a mutual friend of theirs who is

31 McKenna, *Conversation and Responsibility*, 70.

Hispanic. Daphne's failure to invite Leslie gains its meaning in part by the background assumption that they typically go to lunch together, and in part by the meaning she assigns to Daphne's remark.

One more point before proceeding: blaming is taken to have a characteristic force, a sting that is liable to harm the one blamed. How do the elements of the conversational model figure in an account of that harm? Put simply, according to the conversational model, *how* does blame hurt? Drawing upon Joel Feinberg's notion of welfare interests, and assuming that we can understand harms as setbacks to interests, I find three distinctive harms that blaming is liable to cause when we understand blame to have the conversational features I contend it has.³² First, it impedes a blamed person's ability to exercise her capacities to engage in normal social intercourse. Why? Part of the expressive and communicative elements invoked in "conversing" by blaming involve our inclination to withdraw from or instead revise otherwise normal social relations with the one blamed. Second, a person's freedom to live her personal life as she wishes is impeded. Demands that one respond to another's charge of wrongdoing, that she apologize or account for herself, are liable to interfere with living her personal life as she pleases. Third, these sorts of demands and revisions to the expected interpersonal relations among others are emotionally taxing; they affect, or at least are liable to affect, one's emotional stability. So, in summary, on the conversational model, blame is liable to cause three distinctive sorts of harms by threatening the following welfare interests: the ability to engage in normal social intercourse, the freedom to live one's personal life as one wishes, and the preservation of emotional stability.

5. THE DIFFERENCE BETWEEN BLAME AND PUNISHMENT

The preceding details offer enough resources to fit the conversational theory of moral responsibility for a basic-desert thesis for blame. Before turning to that topic, however, I pause to note the way in which blame differs from punishment according to the conversation theory. This is not a minor point. Some philosophers treat blame as informal punishment.³³ In doing so they fix on the idea that blame is a species of sanction. But this gives added reasons for critics of desert-based theories of moral responsibility to damn claims of basically deserved blame by way of arguing against desert-based theories of punishment—especially full-blown retributivist theories. To the extent that blame differs from punishment, it might well be that the features of punishment that critics of retribu-

32 Feinberg, *Harm to Others*, 37–38; McKenna, "Directed Blame and Conversation," 134–41.

33 E.g., Feinberg, *Doing and Deserving*.

tivist theories find especially objectionable are lacking from blame's nature, and by extension, the norms governing its application.

On the conversational theory, blame is a stage in a dynamic, unfolding conversation (conversational analog), and it plays a role in that exchange at an intermediary stage. At the first stage, Moral Contribution, a morally responsible agent performs an act that has agent meaning of a sort that invites the defeasible thought that she is blameworthy. At the second stage, Moral Address, another holding her morally responsible addresses her and expresses her resentment or moral indignation. Here is the blaming stage. Then at a third stage, Moral Account, the blamed agent offers an account, a reply to the blame, in the form of an apology, an excuse, a defiant admission of guilt, and so on. On my view, at least in typical cases, punishment is yet a *further* stage in the unfolding conversation. Its conversational meaning is a response not just to the wrongdoing that instigated the exchange, but also to the account offered at the third stage, Moral Account, wherein the blamed party replies to those blaming her. Punishment's meaning encompasses more than does blame's. (This is analogous to the way conviction and sentencing are treated in the law.)

Here, now, are two ways that on my view blame differs from punishment. First, the harms to which a wrongdoer is exposed in our blaming practices are limited to the range of welfare interests delineated in the preceding section. While this range of harms is after all substantial, and can "really hurt" as the saying goes, there is nevertheless an upper limit on such harms that is lacking in the case of punishment, which can include physical harms, confinement, monetary sanctions, obligations for community service, probationary periods, and formal means of banishment from organizations. Second, and perhaps more important, *by its very nature* punishment involves the intention to harm. But at least on the conversational model, blame need not. Blaming involves expressing and communicating one's regard for a blamed party's quality of will. It involves the conversational expectation and presumed burden of a reply from the blamed party. But its aim *need* not be to harm; its aim need only be to converse, to engage, to demand, and so on. Naturally, it is reasonable to think that competent moral agents engaged in blaming practices will understand that their blaming is liable to harm the one blamed. And so they are likely to harm knowingly. But what we do knowingly is much wider than what we do intentionally. Hence, punishment essentially involves the intention to harm as a means of sanction; blame does not.

6. FITTING THE CONVERSATIONAL THEORY FOR A BASIC-DESERT THESIS

In fitting the conversational theory of moral responsibility for a basic-desert the-

sis, I aim to avoid opposing charges from two different critics. There are those who would reject a basic-desert thesis for moral responsibility because basically deserved blame involves a host of objectionable features that it shares with basically deserved punishment. Then there are those who would reject the conversational theory of moral responsibility if it fell shy of a theory that included basic desert for blame.

6.1. *Resisting the Charge of Overcommitting*

As for the critics who are against a basic-desert theory of moral responsibility because of its unsavory alliance with basically deserved punishment, what features of punishment might make it so objectionable? There are several potential candidates. One has to do with responding to a wrongful harm with a like harm: an eye for an eye. If, as the objection goes, justice is unpacked in terms of basic desert, and this is what desert comes to, justice commands acts of barbarism. There are, furthermore, worries about proportionality, of responding to a wrongful harm with a harm of the same degree. So those who torture and maim must somehow be caused harm proportionate to what they caused? Who is to shoulder this burden? Then there is the thought that justice would command intending to harm the one to be punished. Is this bloodlust? Vengeance? Even if neither, is it not just cruel? An extreme view would have it that to account properly for true moral responsibility one must be able to make sense of a blameworthy person's not just basically deserving blame, but more drastically, as Galen Strawson would contend, the intelligibility of deserving eternal punishment in hell.³⁴

If any of the above reasons are what motivate some to reject a basic-desert thesis for punishment, none translate to the activity of directed blaming on the conversational model. If blaming functions on analogy with a conversational response to one who initiated a conversation, the proper metric is to be understood in terms of the intelligibility or meaningfulness of a reply that extends the conversation. It is not a matter of "saying" to the one who introduced the initial meaningful contribution just what she said back in reply. It is a matter of responding in a way that fits the salience of that contribution in a manner that gives expression to the blamer's demands, expectations, and so forth. Here also it is hard to see how this can be put in terms of proportionality of *harm*. It is better to think of any sort of proportionality in terms of the severity of the response being adequate to convey meaningfully the blamer's disapproval, her hurt feelings, and so forth. Also, as noted in the previous section, worries about the extreme severity of deserved harm are out of place. The range of harms the activity of blaming can cause according to the conversational model are limited

34 Strawson, "The Impossibility of Moral Responsibility."

to the range of welfare interests identified above. Finally, as noted in the previous section, unlike punishment, on the conversational model blaming does not *require* the intention to harm. So charges of being essentially rooted in vengeance are unfounded. All of these considerations suggest that the pertinent harms in blaming should not be indicted with the harms that are taken to flow from basically deserved punishment.

One serious philosophical worry about the very nature of basically deserved blame arises from the idea that harm of any sort could ever amount to an intrinsic good. All harms, so the objection might go, only aggregate along the negative evaluative dimension of badness, negatively adding to any harm already done. Here reasonable minds can differ, and the best way to illustrate this is in terms of the claims of cross-world difference noted above (section 3). But now we can do so, not in terms of a case of punishment, but in terms of a case of blame. Consider a case in which one performs a blameworthy act, say, making a hurtful, racist remark. Now consider two worlds, one in which the wrongdoer is in no way harmed in response to her remark. There is just the harm she causes, and no harm that accrues to her. Then consider another in which the wrongdoer is harmed in response to her wrongful harm in *only* the following ways, and only assuming that some sense of proportionality in the response is achieved: by others blaming her, her ability to engage in some spectrum of social intercourse hampered, her freedom to live her personal life as she wishes disrupted, and her emotional stability unsettled. Hold in mind, in imagining such a case, that the degree of these harms might also be fairly minimal and so would fall far shy of anything like extreme suffering. I fully acknowledge that some are likely to claim that indeed the first world is a better world. Adding one harm to another, they might argue, only increases the amount of intrinsic badness in the world; no intrinsic good can come from the addition. But I disagree. There is something fitting about a world in which a blameworthy wrongdoer is made worse off *in just the limited ways identified here*. That world is a better world than one where the wrongdoer is in no way harmed. Why? Well, one might say, because she deserves it.

6.2. Resisting the Charge of Normative Inadequacy

What of those who would reject the conversational theory because it falls shy of a basic-desert thesis? There is a fair complaint about the bare bones of the theory as set out thus far—that is, when it is not explicitly fitted for a basic-desert thesis. The conversational features of the theory unpack claims of appropriateness in a particular way. So they aid in giving some informative content to the claim that blaming is an appropriate response to one who is blameworthy. Conversational meaningfulness or intelligibility is a thick, informative notion. Indeed, some

who are foes of a desert thesis of any sort might point to the norm of intelligibility as a way to account for the fittingness of blame in the absence of any claim of desert. But the problem is that intelligibility or conversational meaningfulness is just not the right sort of normative warrant to engage in a practice whose nature involves harming those at whom it is directed. Some *further* normative warrant is needed, and so it seems the resources of the conversational theory cannot by themselves justify the harms in blaming, even if the harms that are identified are harms that flow from the conversational nature of the activities associated with directed blaming.

This seems right. The conversational theory of moral responsibility needs supplementing with normative resources that reach outside the conversational elements of the theory. My own view is that this can be done with resources that do not require commitment to basic desert.³⁵ One can do so by appeal to exclusively nondesert-based resources.³⁶ Instead, one can do so by appeal to nonbasic desert resources.³⁷ Nevertheless, as I now argue, one can also supplement the conversational theory by way of a basic-desert thesis. The proposal is simple: Take the harms I have identified in directed blaming on the conversational theory. Now add the claim that one who is blameworthy deserves just this limited range of harms, in the basic sense of desert. I offer only two points to develop this proposal.

First, consider the norm of conversational intelligibility or meaningfulness as a fitting response to a meaningful contribution—the contribution being the presumed agent meaning of the blameworthy agent’s act. The conversational theory offers an illuminating way in which the elements of a desert-base can be said to fit in a case-specific manner the deserved response. Recall that early on (section 1) I noted that it is especially difficult to state the way in which a particular deserved blaming response could, in a case-specific fashion, be fitted for a particular desert-base. In this respect, the conversational theory offers an elegant way to capture this. It is the sort of fittingness that uniquely pairs a particular conversational reply intelligibly for a prior particular meaningful contribution. In this respect the current basic-desert thesis for blame appears to avoid rather easily a (perhaps surmountable) problem for theories of punishment. How so? It is a truism to claim that, on a retributivist theory of punishment, the punishment should fit the crime. And this normative requirement of fit does seem to be motivated by the thought that this is a matter of what one who is to be punished

35 McKenna, *Conversation and Responsibility*.

36 À la the work of a philosopher like Scanlon, *What We Owe to Each Other*.

37 À la the work of someone like Lenman, “Compatibilism and Contractualism”; Wallace, *Responsibility and the Moral Sentiments*; or Vargas, *Building Better Beings*.

really deserves. But with punishment, it is often hard to capture this. A person imprisoned for armed robbery might get five years, and so might a rapist or a corrupt banker. Where is the special fittingness between punishment and crime? Of course, this is not to say that properly justified forms of punishment could not be so tailored, but it stretches the imagination to think of how this could be so. On the conversational theory of moral responsibility, there is an elegant way of conceiving of the special fittingness relation between an agent's blameworthy act and the blame she deserves.

Second, consider a principle like D_2 , which I offered as a way to understand a basic desert thesis for blame:

D_2 : Because it is intrinsically good to harm by blaming one who is blameworthy for a morally wrong act, there is a reason that favors doing so.

It is easy to see how it might be altered to fit the conversational theory, as follows:

D_2^* : Because it is intrinsically good to harm by blaming one who is blameworthy for a morally wrong act (where the harms in blaming are limited just to those identified on the conversational theory), there is a reason that favors doing so.

I offer D_2^* as an initial formulation.

Elsewhere, I have offered three observations to establish the claim of goodness as captured in thought experiments like those involved in the cross-world comparisons offered above (section 4).³⁸ These goods are meant to function as markers for or evidence of the good referenced in a principle like D_2^* . Here are three goods that can be cited when someone asks, "Why are the particular harms of blaming good?" First, it is good for the blameworthy agent that she is harmed in the ways unique to directed blame on the conversational theory. This is because the potential harms to which she is exposed, such as the ability to engage in normal social relations with others, are harms to her only insofar as she is committed to membership in the moral community. Her liability to such harm is an *expression* of her being so committed. That is good. Second, it is good for the one blaming insofar as one's blaming is motivated by and an expression of one's commitment to morality. That too is good. Third, and finally, the relationship between blamer and blamed in the practice of blaming is itself part of an activity whose aim is to ameliorate and sustain the bonds of moral community.³⁹ That also is good. (I defend these claims below.)

38 McKenna, *Conversation and Responsibility*, 167–70.

39 See Christopher Bennett's development of a basic desert thesis ("The Varieties of Retributive Experience"). Bennett makes the aim of our blaming practices moral reintegration, and

7. A SERIOUS CHALLENGE: ARE THE GOODS AT ISSUE
REALLY SUITED FOR BASIC DESERT?

Note that D_2^* is formulated in terms of intrinsic goodness. As such, it is not well suited for the conversational theory. Consider the nature of the harm that, I contend, is good. On reflection, it is a distortion to describe its (disputed) value as intrinsic. Why? The nature of the thing that is claimed to be good, the harm issuing from the conversational dimension of the blaming activity, gains its nature as a communicative and conversational response only by virtue of its relation to a collection of practices and norms against which the blaming activity can have a salience. Furthermore, insofar as its status is part of a stage in an unfolding conversation (or analog to one), it depends on what transpired prior to it as a meaningful contribution, and expectations about what might unfold after it, as felicitous as in contrast with infelicitous responses, for instance. So it seems that its status as good is, after all, *extrinsic*, not intrinsic. Note, furthermore, that the goods I offered as evidence for the goodness of the harm in blaming were all in some way characterized in terms other than those internal, intrinsic features of the harm. All were in some way *extrinsic*. I drew attention to the agent's commitment to moral community, or the blamer's commitment to morality, or the way the interaction between blamed agent and blamer involves an activity aiming to sustain the bonds of moral community.

If what is required of a basic-desert thesis for blame is that the good in the harm of blaming is intrinsically good in a strict sense, then I have after all fallen shy of a basic-desert thesis. Derk Pereboom, for instance, often writes of a basic-desert thesis for blame in such a way that if an agent is blameworthy for an act in the basic-desert-entailing sense, then she deserves blame *just because* she so acted.⁴⁰ Here, the "just because" invites the reading that the ground for the desert, found within the desert-base, and any claim of goodness entailed by what is deserved, cannot reach beyond the mere fact of the agent's so acting. Pereboom, it seems, requires that any good implicit in judgments of deserved blame has to be intrinsic in the strict sense.

In response, I propose a more liberal view of the value of blame. In her influential paper, "Two Distinctions in Goodness," Christine Korsgaard points out that the distinction between intrinsic and extrinsic goodness differs from the distinction between noninstrumental and instrumental goodness.⁴¹ Something

treats this as something that is good, and not merely for consequentialist reasons. Clarke, "Some Theses on Desert," draws upon a similar point.

40 E.g., Pereboom, "Free Will Skepticism, Blame, and Obligation," 189.

41 Korsgaard, "Two Distinctions in Goodness."

might be noninstrumentally good since its goodness is not in the service of another thing, and yet its goodness might be dependent upon its extrinsic relation to other things. A simple example might be the beauty of a rare flower, whose beauty would not be so precious if flowers of that kind were commonplace. Another, I have suggested, is a musician's contribution to a musical piece, where the contribution gives meaning to but also acquires meaning from the accompanying instruments. My example is the drum work by Joe "Philly" Jones on John Coltrane's album *Blue Train*.⁴²

Given this distinction, in earlier work I suggested, but did not develop, a basic-desert thesis in terms of noninstrumental value.⁴³ Here is a revision to D_2^* revised to accommodate this more permissive conception of basically deserved blame:

D_3 : Because it is noninstrumentally good to harm by blaming one who is blameworthy for a morally wrong act (where the harms in blaming are limited just to those identified on the conversational theory), there is a reason that favors doing so.

Does my concession that the pertinent good is merely extrinsic and not intrinsic weaken my thesis? Is D_3 not exposed to the indictment that the particular extrinsic facts that would ground a claim of goodness are also facts that show the good to be not merely extrinsic but, contrary to my contention, instrumental? In short, the charge might go: It is all well and good to distinguish between extrinsic and instrumental goodness. But drawing the distinction does not mean you can carry it off in this case. If the goodness of the harm you identify in blaming gains all of its value exclusively from serving the elements extrinsic to it, then it is not merely extrinsically good; if good at all, it is only instrumentally good.

So it seems I have two burdens. One is to resist the thesis that anything shy of a strict claim of intrinsic goodness falls short of a proper basic-desert thesis. Recall, this arises from Pereboom's "just because" formulation, which seems to limit the thesis to intrinsic rather than any sort of extrinsic but noninstrumental good. On this view, while D_2^* might be a serious contender, D_3 is not, and all the conversational theory offers is something like D_3 . Another, assuming the first burden can be met, is to resist the worry that the goods I have identified as extrinsic but noninstrumental are upon examination merely instrumental goods and nothing more.

42 McKenna, *Conversation and Responsibility*, 170.

43 McKenna, *Conversation and Responsibility*, 123–24n18.

8. MY REPLY: DEFENDING THE ROLE OF
EXTRINSIC AND NONINSTRUMENTAL GOODNESS

Consider the first burden. Citing a good that is an extrinsic but noninstrumental good located in the harm of blaming would be enough to foreclose the worry that the good only gains its value by serving a consequentialist or contractualist end. Indeed, when Pereboom makes these claims about intrinsic goodness, he almost invariably clarifies by writing of the sense in which it differs from goods that could be justified by consequentialism or contractualism.⁴⁴ His most immediate worry seems to be that there is some more general moral theory or more general moral principles providing the justification for blaming, and any goodness in blaming gains its value from that theory or those principles. The extrinsic goodness figuring in the proposed conversational model is not meant to serve an instrumental purpose in relation to other general moral theories or principles.

Furthermore, consider use of the expression “just because” when Pereboom claims that, according to a basic-desert thesis, an agent would deserve blame just because she so acted. It is not clear that we must understand the “just-because” relation so that it rules out the relevant value of the blame being extrinsically related to other things. It might function like an indexical, picking out what can be added, holding fixed other features of a system or set of relations that would then give something a certain value. Granting, for example, that there are so few flowers of a certain sort, and holding fixed other standards of beauty viewers might take to evaluating flowers, with all that in place, one might think a flower, growing alone in an open field is beautiful *just because* it is located there. This might be so for the goodness found in the activity of blaming. Under the assumption that the practices and norms set context and salience for a blaming response, and they provide the constitutive resources to give the response meaning, one might say, adding the response of blaming is fitting—and deserved—just because the agent so acted.

Now consider the second burden. In assessing an earlier formulation of my proposal, Pereboom argued that the value in the communicative and conversational dimensions of the blaming activity, and the harms that attend it, are limited to instrumental goods.⁴⁵ For instance, I cite the good in the blameworthy agent’s commitment to membership in the moral community, claiming that it is noninstrumental. Pereboom replies:

On the basic desert view, it is good that the blameworthy agent is harmed

44 E.g., Pereboom, “Free Will Skepticism, Blame, and Obligation,” 189.

45 Pereboom, “Free Will Skepticism, Blame, and Obligation,” 194–96.

in the ways indicated just because he has knowingly done wrong, and in the context of the debate, this is just what it is for such a harm to be a noninstrumental good. Harm aimed at the good of membership in a moral community would, by contrast, be instrumental, since the harm is not envisaged as good in itself but instead as serving the good of such membership.⁴⁶

Then, in response to my contention that the harm in blaming encompasses a range of noninstrumental goods whose status is dependent upon its relation to others and to sets of practices (and thus is good extrinsically), he writes:

While it is plausible that certain kinds of obvious goods, such as mental and physical health, are partially constitutive of the noninstrumental good such as human flourishing, it is at least typically less credible that harms—as harms—are partially constitutive of noninstrumental goods, and for this reason count as noninstrumental goods themselves. Vaccination may be a prerequisite of physical health, and health constitutive of flourishing, but it is not plausible that the pain of vaccination is constitutive of flourishing, by contrast with being instrumentally required for it.⁴⁷

Pereboom's criticism crystallizes the sort of objection I indicated above: granted, there are noninstrumental goods whose status is extrinsic, but the harms I have identified in the practice of blaming according to the conversational model are not among them.

I offer two points in reply. First, as I have noted previously, the goods that I contend are at issue might well be of a "mixed" variety, and so might thus serve some instrumental purposes.⁴⁸ But they might *also* have noninstrumental value. Crucially, something can be of value because of the aim it takes on or its role in a practice with a certain *telos*—such as aiming at preserving the bonds of moral community—and have that value regardless of whether it actually helps to attain that end. Its value is in its commitment to that *telos*, and (not just) in its instrumental efficacy of achieving the desired end.

Second, recall the cross-world thought experiments I proposed to make sense of how one harm in response to another could constitute a good. Of course, I had originally put that in terms of capturing a sense of intrinsic value. But it can be employed to help establish similar claims about extrinsic, noninstrumental value. The cross-world comparisons just have to include in one world,

46 Pereboom, "Free Will Skepticism, Blame, and Obligation," 195.

47 Pereboom, "Free Will Skepticism, Blame, and Obligation," 196.

48 McKenna, *Conversation and Responsibility*, 170.

and exclude in another, the extrinsic relations to the pertinent harmful activities that are alleged to issue in the good-making feature. To be clear, what is it that is a candidate for a noninstrumental good? It is the harm itself as it contributes to the blaming activity, an activity that is in response to, and so is extrinsically dependent upon, an antecedent blameworthy act. Bearing this in mind, Pereboom contends that it is less credible that harms can be partially constitutive of a non-instrumental good, as in comparison with certain benefits that can be partially constitutive of a noninstrumental good. As quoted above, he illustrates with a simple case of the pain of vaccination (a harm), which is only instrumentally good in the service of health and human flourishing. This is a powerful challenge, but note that it turns primarily on the force of the sort of example he enlists.

Here is different sort of case. Consider grief as a response to the loss of the parents one loves, or in response to losing a sibling early in that sibling's life. Grief as part of the good of living a flourishing life might be seen as a noninstrumental good insofar as it is bound up with accepting the realities of our loving relations and our finitude. Grief is certainly a harm, but sometimes it is also a good, a noninstrumental good connected with accepting our humanity. Consequences aside, a life without grief in response to losing the ones one loves would be worse than one where, in the face of loss, one underwent an appropriate period of psychological pain and mourning. One who would not mourn the loss of their mother, for instance, is one whose life is in some way impoverished. Or at any rate, even if one disagrees, one can see the point of this sort of claim. The harm in blaming, a basic-desert theorist might argue, *is like that*. It is intimately connected with commitments and modes of life within a moral community whose aim is itself good—and good not merely as an instrument for something else. Crucial to this test case for resisting Pereboom is the idea that grief might very well contain an element that is noninstrumentally or intrinsically bad while nevertheless contributing to a whole that is intrinsically or noninstrumentally good.

I take the preceding case of grief to be adequate to respond to Dana Nelkin's challenge to my view.⁴⁹ However, exploring how she might resist me will help sharpen my thesis. Like Pereboom, Nelkin also wishes to resist my claim that pertinent harms are good. Nelkin contends that the goods I identify in blaming, such as the care a blamed person might have for others' regard, can indeed be noninstrumentally good. And it might well be, she contends, that the harms that attend them come with these goods as manifestations of them, but it just does not follow from that that the goods I identify are good *in virtue* of these harms. Hence, it is, for instance, the caring about others that is, as she puts it, fundamentally good and not the harm that, she grants, may be non-contingently related to

49 Nelkin, "Moral Responsibility, Conversation, and Desert."

it. This is an excellent way to capture what is at stake. It poses a serious challenge to my contention that there are these goods-as-harms identified in the practices of blaming. In the case of grief, the challenge might go, what is fundamentally good is one's caring for family or friends, and a life that involves the attendant intimacies that promote one's flourishing. The pain or harm of loss might be non-contingently related to this thing of fundamental value, given the fragilities of our human lives, but it simply does not follow that it is *itself* good.

Now why take grief to be the right sort of case to resist Nelkin's insightful challenge? Rather, does she not show precisely why my appeal to a case like grief falls short? I grant, on its face, it seems to. Before offering a direct reply, note the following qualification: it is consistent with my view that it would even be a better world if it were the case that one could enjoy and celebrate the goods of family and friendship without the liability to grief. This might fallaciously lead one to think that in a world where there is the good of friendship and familial love that *is* accompanied by the pain of loss, that the pain or loss itself could not be noninstrumentally good in any way as a mode of contributing to one's caring and loving. But this is just not true. Were we differently and perhaps better equipped beings, we might not be exposed to these vulnerabilities. But given that we are so exposed, there is the question of whether the harm attendant with grief due to caring can be a noninstrumental good.

I turn now to my direct reply. Critics such as Nelkin appeal to an enticing analytic move by distinguishing the dimension of, say, grief that is appealing or seemingly good-making, which is the caring, and then separating that from the dimension that is unappealing and so seemingly bad-making, which is the distinctive pain of loss. This allows Nelkin to contend that any goodness in grieving (or blaming) is a goodness in virtue of just that dimension, the caring dimension, not the other dimension, the one to which we are averse, the painful ingredient. But I reject this analytic move, this prying apart of the good-making feature as distinct from the alleged bad-making feature. Of course, it will not help merely to contend that, in certain cases, the alleged bad-making feature is necessary or non-contingently related to the good-making feature. Nelkin grants that. Instead, the burden here is to show that the relevant harm actually *contributes* to the goodness. And in response, what I wish to say is that, as a *distinct thing*, it would not. Merely as harm, it would not contribute to any goodness. But what a basic-desert theorist can say instead is that it is mistaken to infer from the fact that we can identify these different dimensions of grief or blaming that these can be understood as fully distinct ingredients—like separate ingredients added together to bake a cake—rather than as a unity whose whole has a value that is not analytically decomposable in this way. They are not like separate parts that just “add up.”

I suggested above that the liability to certain harms can be an expression of one's cares. Furthermore, the propensity of others to blame and in doing so harm is a way of registering the concern of the community of blamers. One way to understand this is that such expression is itself, accompanied by these distinctive harms, noninstrumentally good as an *organic unity*. It would not be good as that particular expression of one's cares absent its expression as a form of harm or suffering. When considering the nature of value as it bears upon the problem of evil, Marilyn Adams usefully draws on a distinction introduced by G. E. Moore and Roderick Chisholm.⁵⁰ There is the evil that is balanced by a greater good, but there is also evil that is *defeated* by what is good, and in this defeating, one finds a kind of goodness. So too with grief. There is pain in grief on particular occasions, and so a kind of harm, even something that, it can be granted, is intrinsically bad, but as an expression of *this* way of caring for *this* person who suffered *this* loss, it is a "bad" or an evil that is not merely outweighed but defeated. The key distinction between merely outweighing and defeating is that in outweighing a good can be considered as making a positive contribution whose value is not dependent on the bad. The thing of disvalue could, so to speak, be subtracted and the good still stand.⁵¹ In defeating, the good gains its significance as a good by being a response to and finding goodness built from the thing of disvalue. So, for instance, in expressing one's affections, in expressing one's love when undergoing a period of grief, that pain in loss counts in the context of that set of extrinsic relations as something that is good given the good of the grieving response. So too, I would say, for the harms I have identified in the activities of blaming according to the conversational model. Yes it is, for instance, a harm for the one blamed that she suffer the setback of others distancing themselves from her, but as this harm for her is also an expression of her concern for others, it likewise counts as a good that she is harmed in this way.

9. CONCLUSION

I offer the preceding as an account of basically deserved blame. On my proposal, what a blameworthy person deserves in a basic sense is a directed blaming response of a conversational nature, one that engages the blamed person in such a way that she registers the moral demands and concerns of those engaging her. Moreover, the harms that she is alleged to deserve—insofar as she deserve blame

50 Marilyn Adams (*Horrendous Evils and the Goodness of God*, 55) usefully draws on a distinction introduced by Moore (*Principia Ethica*) and Chisholm ("The Defeat of Good and Evil").

51 Cf. Moore, *Principia Ethica*, 29.

and not also something more like punishment—are *exhausted* in the activities of blaming her, and of her registering them as such. These are social harms regarding the blamed person's relations with others. When expressed in a fitting fashion, say by pertinent conversational standards of intelligibility or meaningfulness, they engage the person as a moral agent and as a member of the moral community, or at least a potential member.

In developing this view, I have committed to the axiological thesis that when deserved, the pertinent harms of blaming are noninstrumentally good. Many, I suspect, will reject this ingredient as a veiled form of barbarism justifying vengeance and inclinations toward brutality. Indeed, as noted above (section 1), when considering a value-based conception of blame, Scanlon was emphatic in distancing himself from any such view about the goodness of causing the blameworthy to suffer.⁵² So too for Wallace.⁵³ Maybe after all this is the correct view, and so maybe one should, as Scanlon did in later work, opt for an exclusively deontic version of a desert thesis for blame.⁵⁴ But note two considerations before rejecting an axiological view of the sort I have defended.

First, one should bear firmly in mind that the axiological claims of noninstrumental goodness involved in the desert thesis presented here are quite limited. What does a blameworthy person deserve on this view, and what would be noninstrumentally good for her to receive? No more than what is involved in that person having a proportionately pained response to others altering their interpersonal relations with her—and altering them as would befit their communicating to her their moral demands and concerns from a place of proportionate moral anger. *That is all.* No aim of writhing on the floor or the demand for the wearing of hair shirts is part of the mix.

Second, those favoring an exclusively deontic version face burdens of their own. As I argued above, if a strong version is ruled out for deserved blame, what is left is a weak version permitting but not requiring blaming the blameworthy. But merely permitting that the blameworthy be harmed by blaming supplies no reason why one blaming *should* harm them. So how do the exclusively deontic theorists avoid the charge that they are the ones who permit gratuitous harm under the banner of desert? They cannot really say it is not gratuitous *because* the person deserves it, since the deserving itself is—by hypothesis—no reason to favor treating the person in that way. So there is some reason to worry that it is after all the exclusively deontic versions of a basic-desert thesis that might just help to conceal barbarity and vengeance.

52 Scanlon, *What We Owe to Each Other*.

53 Wallace, *Responsibility and the Moral Sentiments*.

54 Scanlon, *Moral Dimensions*.

I close with one final point about whether anyone deserves blame in the basic sense specified here. In this paper, I have only attempted to articulate a theory of what basically deserved blame is. I take it to be an open philosophical question whether anyone deserves it. I assume that if no one has free will, no one deserves to be blamed in this basic sense of desert. Nothing in the preceding discussion was intended to settle the freewill question. As a compatibilist, I do think that most people possess free will and that most of the time they act freely. However, I grant this is a contestable view. As result, so is the question of whether anyone deserves blame in a basic sense of desert. Nevertheless, one point I hope to have established has to do with *what* a blameworthy person would deserve if she did have free will, and what would be good about it. Some philosophers, such as Galen Strawson, have argued that the freedom at issue in the freewill debate is the type required to make intelligible deserved eternal suffering in hell.⁵⁵ In my view, this helps to set the intuitive bar for the sort of freedom needed so high that it quickly becomes clear that nothing metaphysically possible for finite beings like us could hit that bar, which is exactly the conclusion Galen Strawson himself draws. But if the theory of deserved blame offered above is anywhere within the vicinity of correct, the freedom that would be needed to deserve blame would only help ground a culpable person's being subject to the limited interpersonal social harms identified here. That still might require a fairly robust freedom, but at least it seems to be within the reach of mere mortals like us.⁵⁶

University of Arizona
msmckenna@email.arizona.edu

55 Strawson, "The Impossibility of Moral Responsibility."

56 I have profited from two excellent sets of referee reports from the *Journal of Ethics and Social Philosophy*, as well as two excellent sets of referee reports from *Ethics*. I am grateful to each of these referees for their careful attention. I would like to thank Dana Nelkin, Derk Pereboom, Guido Pincione, Carolina Sartorio, and David Shoemaker for helpful written comments on an earlier draft of this paper. I presented earlier versions of this paper to the Philosophy Department at the University of Calgary in fall 2013, and in the summer of the same year to a workshop in Park City, Utah, sponsored by the University of Arizona Center for the Philosophy of Freedom. I also presented yet a further revised version at the Gothenburg Responsibility Conference in Gothenburg, Sweden, in August 2016. For helpful comments at the Calgary session, I would like to thank Justin Caouette, Randolph Clarke, Ishtiyaque Haji, Daniel Haas, Noa Latham, Mark Migotti, and Angela Smith. For helpful comments at the Park City session, I would like to thank Elijah Milgram, Carmen Pavel, Guido Pincione, Daniel Russell, David Schmidtz, and Steve Wall. For helpful comments at the Gothenburg session, I would like to thank Santiago Amaya, Joe Campbell, Randolph Clarke, Derk Pereboom, Ingmar Persson, Paul Russell, David Shoemaker, Saul Smilansky, Helen Steward, Matt Talbert, and Kadri Vihvelin.

REFERENCES

- Adams, Marilyn McCord. *Horrendous Evils and the Goodness of God*. Ithaca, NY: Cornell University Press, 1999.
- Bennett, Christopher. "The Varieties of Retributive Experience." *Philosophical Quarterly* 52, no. 207 (April 2002): 145–63.
- Chisholm, Roderick. "The Defeat of Good and Evil." *Proceedings of the American Philosophical Association* 42 (1968–69): 21–38.
- Clarke, Randolph. "Some Theses on Desert." *Philosophical Explorations* 16, no. 2 (2013): 153–64.
- Feinberg, Joel. *Doing and Deserving: Essays in the Theory of Responsibility*. Princeton: Princeton University Press, 1970.
- . *Harm to Others: The Moral Limits of the Criminal Law*. New York: Oxford University Press, 1986.
- Fischer, John Martin, and Mark Ravizza. *Responsibility and Control: An Essay on Moral Responsibility*. Cambridge: Cambridge University Press, 1998.
- Hieronymi, Pamela. "The Force and Fairness of Blame." *Philosophical Perspectives* 18, no. 1 (December 2004): 115–48.
- Korsgaard, Christine. "Two Distinctions in Goodness." *Philosophical Review* 92, no. 2 (April 1983): 169–95.
- Lenman, James. "Compatibilism and Contractualism: The Possibility of Moral Responsibility." *Ethics* 117, no. 1 (October 2006): 7–31.
- Macnamara, Coleen. "Reactive Attitudes as Communicative Entities." *Philosophy and Phenomenological Research* 90, no. 3 (May 2015): 546–69.
- McKenna, Michael. *Conversation and Responsibility*. New York: Oxford University Press, 2012.
- . "Directed Blame and Conversation." In *Blame: Its Nature and Norms*, edited by D. Justin Coates and Neal A. Tognazzini, 119–40. New York: Oxford University Press, 2013.
- McNaughton, David. *Moral Vision: An Introduction to Ethics*. New York: Blackwell, 1988.
- Mele, Alfred. *Free Will and Luck*. New York: Oxford University Press, 2006.
- Moore, G. E. *Principia Ethica*. New York: Cambridge University Press, 1903.
- Nelkin, Dana. "Moral Responsibility, Conversation, and Desert: Comments on Michael McKenna's *Conversation and Responsibility*." *Philosophical Studies* 171, no. 1 (October 2014): 63–72.
- Pereboom, Derk. *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press, 2014.
- . "Free Will Skepticism, Blame, and Obligation." In *Blame: Its Nature and*

- Norms*, edited by D. Justin Coates and Neal A. Tognazzini, 189–206. New York: Oxford University Press, 2013.
- . *Living Without Free Will*. Cambridge: Cambridge University Press, 2001.
- Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.
- Sartorio, Carolina. *Causation and Free Will*. Oxford: Oxford University Press, 2016.
- Scanlon, T.M. “Giving Desert Its Due.” *Philosophical Explorations* 16, no. 2 (2013): 101–16.
- . *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Belknap Press, 2008.
- . *What We Owe to Each Other*. Cambridge, MA: Harvard University Press, 1998.
- Smart, J.J.C. “Free-Will, Praise and Blame.” *Mind* 70, no. 279 (July 1961): 291–306.
- Shoemaker, David. *Responsibility from the Margins*. Oxford: Oxford University Press, 2015.
- Strawson, Galen. “The Impossibility of Moral Responsibility.” *Philosophical Studies* 75, nos. 1/2 (August 1994): 5–24.
- Strawson, Peter. “Freedom and Resentment.” *Proceedings of the British Academy* 48 (1962): 1–25.
- Vargas, Manuel. *Building Better Beings: A Theory of Moral Responsibility*. New York: Oxford University Press, 2013.
- Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press, 1994.
- Watson, Gary. “Responsibility and the Limits of Evil: Variations on a Strawsonian Theme.” In *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, edited by Ferdinand Schoeman, 256–86. Cambridge: Cambridge University Press, 1987.

AGAINST JEFFREY HOWARD ON ENTRAPMENT

Jonathan Stanhope

IMAGINE WICKED BARRY, impressionable Carl, and innocent Leila. Suppose Barry fosters in Carl an intention to kidnap and torture Leila, in a way that does not negate Carl's moral responsibility.¹ (It does not amount to brainwashing, say.) Hence, if Carl actually does kidnap and torture Leila, he wrongs her. And Barry wrongs Leila too, by culpably encouraging Carl.

Plausibly, this completes the catalog of wrongs in our example. But in a recent paper, Jeffrey Howard develops an intriguing argument to the contrary.² He contends that, by inducing Carl to do wrong, *Barry wrongs Carl*. Ironically, Carl shares something with Leila: a complaint against Barry.

The implications of Howard's argument—if correct—are profound. First, entrapment by law enforcers is wrong even if it does not lead to the prosecution or punishment of the entrapped person.³ Second, the wrong of entrapment is, *ceteris paribus*, indistinguishable from the wrong committed by private citizen Barry. Third, governments wrong many of their citizens in entrapment-like ways through policies that make it rational for them to act wrongly. For example, inadequate education, job scarcity, and permissive firearms laws foreseeably encourage some citizens to join criminal gangs. Consequently, our obligations to reform or transform prevailing political and socioeconomic conditions are even stronger than we initially thought.⁴

Howard's animating idea is that inducements to do wrong subvert or interfere with the induced agent's moral capacities.⁵ He believes that we each have a categorical, constraint-imposing duty not to foreseeably increase the likelihood that another agent (culpably) acts wrongly—hereafter DUTY—that is grounded

1 Elaboration of an example in Howard, "Moral Subversion and Structural Entrapment," 29–30.

2 Howard, "Moral Subversion and Structural Entrapment."

3 Nor any other harm, such as reputational damage.

4 Howard, "Moral Subversion and Structural Entrapment," 45–46.

5 Howard, "Moral Subversion and Structural Entrapment," 25.

in a more basic obligation “to respect the first moral power”—hereafter FMP—of that agent.⁶

I accept the existence of the more basic obligation. Nonetheless, whether it entails DUTY, and thereby condemns entrapment and neighboring phenomena, depends on how FMP’s value is grounded. For Howard, the value of FMP is such as to demand a regulative attitude of support for FMP’s *successful* exercise rather than its exercise *per se*.⁷ In what follows, I offer a more nuanced account of the value of FMP, one more attuned to certain liberal thoughts about agency. This undermines Howard’s case for DUTY, and by extension his case against entrapment and its neighbors. (A fact that, I conclude, proves especially significant for the all-things-considered permissibility of entrapment by law enforcers.) Still, in a constructive spirit, I also sketch a different argument for DUTY. Though it does not yet persuade me, I hope it can be developed in a compelling direction.

I

Borrowing from John Rawls, Howard construes moral agency in terms of two moral powers.⁸ In doing so, he notes that the second—the capacity to set, revise, and pursue a conception of one’s rational advantage—has received much greater attention in liberal political philosophy than the first. Whereas the second is central to autonomy-based arguments against paternalism and perfectionism, for instance, the first rarely receives extensive discussion.⁹ Howard’s focus on FMP is a welcome change.

On Howard’s definition, FMP is an “agent’s capacity to reason about right and wrong and to regulate his conduct by the conclusion of his reasoning.”¹⁰ This, quite properly, omits any reference to correctness: FMP can lead the agent to defective conclusions, or—as with *akrasia*—fail to generate right conduct even when the agent has moral knowledge. Yet in outlining what respect for FMP involves, Howard writes that one aspect is “recognizing the power’s existence: the person’s capacity to reason about the demands of morality and live up to those

6 Howard, “Moral Subversion and Structural Entrapment,” 31.

7 Neither I nor Howard presuppose a realist interpretation of moral rightness. As per constructivism, reaching correct moral judgments might consist in, rather than being facilitated by, valid reasoning.

8 Rawls, *Political Liberalism*, 48–54.

9 Howard, “Moral Subversion and Structural Entrapment,” 29–30. Though the first surfaces repeatedly, for example, there are no prolonged discussions of FMP in Freeman, *The Cambridge Companion to Rawls*, or Mandle and Ready, *A Companion to Rawls*.

10 Howard, “Moral Subversion and Structural Entrapment,” 29. Strictly speaking, then, FMP is a bundle of powers, both cognitive and conative.

demands.”¹¹ Not *what the agent thinks are those demands*, but simply those demands, period.

In addition, Howard argues that the second, more demanding component of such respect is a regulative attitude of support for FMP’s *successful* exercise. And this follows from the idea that “the value of the first moral power inheres largely in its successful exercise.”¹² To respect FMP is to respect its proper functioning, given that FMP’s value is grounded in the value of the right conduct it enables. Since foreseeably increasing the likelihood that another agent acts wrongly is incompatible with respecting FMP, it is wrong. (The flip side of this is, of course, DUTY.) Furthermore, it wrongs the agent whose FMP is subverted: *her* moral agency is disrespected.¹³

According to Howard, a wide range of actions count as violations of DUTY.¹⁴ As well as actions that affect conation, such as incitement and temptation, they also include moral-epistemic subversion, where false beliefs about the content of morality or its proper role in practical reasoning are fostered. For example, inculcating the belief that moral reasons are no more decisive or regulative than nonmoral reasons raises the chance that the misled agent will overlook or underweight them.¹⁵

Note that I accept—at least for the sake of argument—Howard’s key assumption that nonmoral reasons cannot override moral reasons. FMP’s deliverances should be the verdict of practical reasoning generally, even when this just means rubber-stamping the deliverances of the second moral power. With Howard, I will not allow for subversions of FMP that respect the agent’s overall practical powers in virtue of giving her overriding reason to act immorally.

II

As stated, my objection is to cashing out the value of FMP in terms of its successful exercise. Howard does not defend this move at length, merely claiming

11 Howard, “Moral Subversion and Structural Entrapment,” 30.

12 Howard, “Moral Subversion and Structural Entrapment,” 30.

13 Howard presumably thinks the degree to which it wrongs the agent is a function of the degree to which it foreseeably increases the probability of her acting wrongly.

14 Importantly, Howard constrains DUTY-violating actions to actions unprotected by an antecedent right. Howard believes that, if it were even *pro tanto* wrong for a person to visit a neighborhood where her skin color elevates the chance of her being assaulted, then it would be appropriate for her to feel regret. Additionally, “fundamental standards of justice” would vary with different probabilities of compliance with them. Both implications are, in Howard’s view, implausible. See “Moral Subversion and Structural Entrapment,” 32.

15 Howard, “Moral Subversion and Structural Entrapment,” 31.

that agents expect—normatively speaking—right conduct from each other.¹⁶ Of course, all going well, FMP does allow agents to meet these expectations. We might even say that it *aims* at meeting them, that it has such a *telos*. In exercising FMP, however, we are trying to discern relevant principles, how to prioritize them, how they interact, and what prescriptions they issue when conjoined with nonmoral facts. We are also trying to integrate these prescriptions into our practical lives so that, in addition to acting as we see right, we find the experience tolerable and hopefully enriching. Aiming to meet normative expectations is aiming to respond to reasons, and is thus characteristically deliberative.

Exercising FMP, then, exemplifies our agency. And it does so whether it leads to right or wrong conduct. Howard's silence on this possibility is odd given that he criticizes paternalism for disrespecting the second moral power. The standard liberal claim about the value of the capacity for a conception of one's good is not that, left to its own devices, the capacity tends to promote one's good.¹⁷ Instead, the claim is that we exemplify agency when we decide for ourselves what lives to lead and how to lead them, even if we decide wrongly. Liberals typically are not skeptical of the existence of an agent's good, nor the suggestion that she might fail to understand or realize it. Rather, they tend to dispute the permissibility of—or place justificatory obstacles in the way of—interfering *as such* with (uncoerced, self-regarding) choices.

Something similar can be said about the value of FMP. Agency is exemplified no less when we reason about how to treat others, and act on that reasoning, than when we reason about our personal fulfilment, and act on that reasoning. True, we might distinguish between personal and moral autonomy, and make it a necessary—and perhaps sufficient—condition of moral autonomy that the agent acts rightly. But that will not stop defective exercises of FMP from exemplifying agency, unless agency just is, or depends on, moral autonomy. Patently, though, it is a demerit for arguments in practical ethics to rely—as Howard does not—on the (presumably) Kantian backstory that would be required. Moreover, even if FMP's value is not connected to agency, it has everything to do with (what we might call) *shmagency*, reflective choice under a different label.¹⁸

16 Howard, "Moral Subversion and Structural Entrapment," 30.

17 In a slogan, "to be able to choose is a good that is independent of the wisdom of what is chosen." See Dworkin, "Paternalism," 72. Whether the value of agency must be justifiable in the qualified terms of public reason is an object of disagreement between liberals.

18 I am shamelessly borrowing the term from Enoch, "Agency, Shmagency."

III

So, it is not promising to exclude from an account of FMP's value the idea of exemplifying agency. At this point, however, it becomes crucial to note that general talk of FMP's value is actually extremely unhelpful. For it is not relevant whether, as Howard believes, the value of FMP is grounded largely in its successful exercise. If FMP demands a regulative attitude of support, rather than being supported when and because it succeeds as a means to a valuable end, this must be because it has final value, value for its own sake.¹⁹ But nothing Howard says indicates that the value of FMP, when successfully exercised, goes beyond its instrumental value. Nor does he clarify how the mere abstract potential for FMP to facilitate right action could be sufficient to confer final value on a *particular agent's* FMP when she exercises it *unsuccessfully*. Why not conclude that the abstract potential for FMP to facilitate *wrong* action confers final *disvalue* on an agent's FMP when she exercises it successfully?²⁰

At a minimum, I think we should believe that the final value of FMP is not wholly grounded in what its successful exercise achieves. In all likelihood, it is not *largely* thus grounded either, given standard liberal thoughts about the value of agency. Indeed, when our axiological picture is sufficiently nuanced, reserving a place for the instrumental value of FMP, I do not feel any pull toward including the goodness of successful exercise in a general account of FMP's final value. (And only slightly more pull when thinking about particular agents' well-functioning FMPS.) Consequently, I am yet to see that the obligation to respect another agent's FMP generates DUTY, and therefore the wrongfulness of entrapment and similar phenomena.

IV

If a regulative obligation to respect FMP would require support for even defective exercises, could that obligation exist? The question suggests a *modus tollens*. Nevertheless, I think we can affirm the basic duty as long as we endorse a fairly minimalist picture of what respect for FMP involves.

Here is a suggestion: it involves providing agents with the fundamental conditions for reasoning to, and acting from, moral conclusions. These include: ad-

19 See Korsgaard, *Creating the Kingdom of Ends*, ch. 9, especially 250.

20 In fact, whether a particular agent's FMP has even instrumental value depends on how effective it is. So, I do not see how to reach a general, impersonal claim about the instrumental value of FMP. Is not that sort of claim just an abstraction from lots of claims about particular agents' FMPS?

equate nutrition and sleep; the absence of avoidable brain injury, indoctrination, and excessive pain; decent secondary-level education; and the political liberties. We already know these things are important and overdetermined by obligation.

Admittedly, another fundamental condition might be a certain sphere of personal choice protected by moral claims against interference with wrongdoing. Certain moral rights to do wrong might well be necessary for the autonomous self-constitution of an integral moral character or identity.²¹ If so, then on some occasions we should stand back when an agent employs her FMP unsuccessfully. Plus, for the protected sphere to be meaningfully wide, it will probably include some choices to wrong other agents: to betray a partner or break a promise, say, even if not to murder and rape. Thus, respect for moral agency might yet be a morally uncomfortable business on my proposal, beyond anodyne requirements like keeping people nourished and not torturing them. Indeed, even if interference with wrongdoing is always all-things-considered permissible, it might be a *pro tanto* wrong that warrants a measure of regret.

V

Suppose Howard remains committed to disconnecting respect for FMP from respect for defective exercises of FMP. While thereby avoiding the question of rights to do wrong, he must shift his grounds for opposing entrapment away from respecting agency. One line might be that acting wrongly is intrinsically bad for an agent, so that to increase the chance of her acting wrongly is to threaten her welfare. However, this claim would have to grapple with the grounds of FMP's final value, since the intrinsic effect of one's wrongdoing on one's welfare surely depends on the nature of agency. Indeed, the current proposal risks denying the content independence about the final value of agency that I have tried to motivate. Furthermore, we should guard against confusing the plausible idea that moral *vices* are bad for an agent with the idea that acting wrongly is bad for her even if it is at odds with her (excellent) character.

A more compelling shift of focus, in my view, would be toward something mentioned earlier: the *telos* of FMP. If we can say that FMP aims at right action, then perhaps we can build a case for the wrongness of interfering with its realization. It could be essential to this argument that right action has final value; after all, interfering with the *telos* of a screwdriver is not intrinsically wrong, since turning screws is only instrumentally valuable. But questions about the final value of FMP could be bypassed, and the standard liberal grounding accepted. On the other hand, given that *the aim* of FMP need not be any particular agent's aim,

21 See, in particular, Herstein, "Defending the Right to Do Wrong," especially 357–61.

this strategy might struggle to justify a directed duty not to entrap, one whose violation would wrong the entrapped person.²²

To retain the idea that FMP is a genuine power whose exercise demands respect, we would have to construe FMP's *telos* in terms of generating right action for the right reasons, or otherwise in a way bringing credit to the agent. If so, then some inducements to do *right* will be condemned, for instance, offers that generate a desire so overwhelming that the agent's judgment is suspended.²³ One challenge would be to explain why interference with a *telos* is the real problem here, instead of suspension of agency. Another challenge, if one accepts the permissibility of pro-moral, subrational "nudges," would be to show how these can evade censure by the current proposal.²⁴

VI

That sketchy proposal aside, my own *telos* here has been to cast doubt on the attempt to derive DUTY from an obligation to respect FMP, and thereby—more importantly—to undermine a recent argument for the intrinsic (and directed) wrongness of ordinary as well as structural entrapment.

If I am right, systematic social injustice is somewhat less wrongful than Howard believes. However, the practical implications of my paper are *far greater* for the conduct of law enforcement than for general matters of social and economic policy. Whether or not Howard's argument is correct, we have overriding reason to change government policies that encourage or allow deprivation. The systematic thwarting of people's life chances already has little if anything to be said for it, and much to be said against it. In law enforcement cases, though, Howard's argument seems more likely to make an all-things-considered deontic difference. And this is for two reasons.

First, there *is* sometimes a *prima facie* case for testing an individual's proclivity toward serious wrongdoing, and seeking to contain it by establishing grounds for conviction. Not least, it can (significantly) reduce the probability of serious harm to (large numbers of) innocent people. Second, the most obvious typical features of entrapment—including manipulation, deception, and the state aiming to restrict substantially the liberty of a citizen—are difficult to view as

22 A point somewhat obscured by Howard's talk of our "quests to be just persons." See Howard, "Moral Subversion and Structural Entrapment," 25.

23 See Frankfurt, "Coercion and Moral Responsibility," 41–42. In the same boat are "Pavlovian" inducements to do right, which risk making an agent dependent on certain cues or rewards in order to act rightly.

24 See Thaler and Sunstein, *Nudge*.

absolutely impermissible. So, in positing a regulative, constraint-imposing duty not to entrap, Howard promised to do much more than throw a defeasible anti-entrapment consideration into the mix. The failure of his argument is an especial loss to opponents of ordinary entrapment—a practice that I am inclined to believe is sometimes permissible and even obligatory in our nasty, dangerous world.

University of Oxford
jonathan.stanhope@mansfield.ox.ac.uk

REFERENCES

- Dworkin, Gerald. "Paternalism." In *Mill's On Liberty: Critical Essays*, 61–82. Oxford: Rowman and Littlefield, 1997.
- Enoch, David. "Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action." *Philosophical Review* 115, no. 2 (April 2006): 169–98.
- Frankfurt, Harry. "Coercion and Moral Responsibility." *The Importance of What We Care About*. New York: Cambridge University Press, 1988.
- Freeman, Samuel, ed. *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press, 2003.
- Herstein, Ori J. "Defending the Right to Do Wrong." *Law and Philosophy* 31, no. 3 (May 2012): 343–65.
- Howard, Jeffrey W. "Moral Subversion and Structural Entrapment." *Journal of Political Philosophy* 24, no. 1 (March 2016): 24–46.
- Korsgaard, Christine M. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, 1996.
- Mandle, Jon, and David A. Ready, eds. *A Companion to Rawls*. Chichester: Wiley-Blackwell, 2014.
- Rawls, John. *Political Liberalism*. New York: Columbia University Press, 2005.
- Thaler, Richard H., and Cass R. Sunstein. *Nudge: Improving Decisions about Health, Wealth and Happiness*. New Haven: Yale University Press, 2008.

DIETZ ON GROUP-BASED REASONS

Magnus Jedenheim Edling

MANY PEOPLE think not only that individuals have reasons to act, but that groups do too.¹ Suppose that they are correct about this. Do the members of a group “inherit” the group’s reason? Alexander Dietz has recently argued that they do so in some circumstances.²

Dietz considers two principles. The first one—which he calls the “Simple Principle”—claims that the members of a group always inherit the group’s reason. The second one—which I call “Dietz’s Principle,” since it is the one Dietz advocates—claims that the members of a group inherit the group’s reason when they cooperate. Although Dietz thinks that the Simple Principle is intuitively appealing, he argues that it has to be rejected because there is a powerful counterexample to it. In this article, I show that there is a powerful counterexample to Dietz’s Principle as well.

I proceed as follows. In sections 1–2, I present the Simple Principle and Dietz’s argument against this principle. In section 3, I introduce Dietz’s Principle and show that it has the intuitively correct implications in the case that is a counterexample to the Simple Principle. In section 4, I turn to my case against Dietz’s Principle. Finally, in section 5, I consider a natural revision of Dietz’s Principle but conclude that it is unsatisfactory.

1. THE SIMPLE PRINCIPLE

The description I gave of the Simple Principle above was incomplete. It is not only concerned with reasons in favor of actions but also with reasons against actions. Here is the complete principle:

The Simple Principle: If a person is a member of a group such that the

- 1 For example, Dietz, “What We Together Ought to Do”; Jackson, “Group Morality”; and Tännsjö, “The Morality of Collective Actions.” Tännsjö does not think that people have reasons to act but is concerned with obligations to act.
- 2 Dietz, “What We Together Ought to Do.”

group has a reason (not) to perform a group action ϕ , then that person has a reason (not) to do her part of ϕ .³

Some comments are in order. First, a group action is a combination of actions. If I perform action A_1 and you perform action A_2 , you and I also perform the combination of actions $\langle A_1, A_2 \rangle$. Furthermore, the combinations of actions a group of people can perform in some circumstances are a function of what individual actions the members of the group can perform in those circumstances. For example, if I can perform action A_1 and action B_1 , and you can perform action A_2 and action B_2 , you and I can together perform $\langle A_1, A_2 \rangle$, $\langle A_1, B_2 \rangle$, $\langle B_1, A_2 \rangle$, $\langle B_1, B_2 \rangle$. There are several views on under what circumstances a combination of actions qualifies as a group action. For the sake of simplicity, I assume that all combinations of actions performed by at least two agents are group actions.

Second, Dietz thinks that the Simple Principle applies to several other moral reasons apart from reasons to make outcomes better; for example, reasons not to harm and reasons to benefit oneself. I am exclusively concerned with reasons to make outcomes better. Third, Dietz calls a reason that is inherited from a group's reason a "group-based" reason. Fourth, Dietz takes the Simple Principle to be an explanatory principle. For example, he takes the fact that a person is a member of a group that has a reason to perform a certain action to explain why that person has a reason to perform a certain action (i.e., his part of the group action in question).

2. AN ARGUMENT AGAINST THE SIMPLE PRINCIPLE

Dietz advances two arguments against the Simple Principle.⁴ For our purposes, it is sufficient that we consider what Dietz takes to be the most important one.⁵ This argument proceeds from the following case, which I call "Impending Disaster." A million lives are at risk and you and I face the following options (our actions are counterfactually independent):

- 3 Dietz, "What We Together Ought to Do," 969. Christopher Woodard has defended a view similar to the Simple Principle in the respect that it imposes no constraint to the effect that the parties should be willing to cooperate on the existence of group-based reasons (*Reasons, Patterns, and Cooperation*).
- 4 Dietz, "What We Together Ought to Do," 968–73.
- 5 Dietz suggests that there is a way around the other argument ("What We Together Ought to Do," 969–70).

		You	
		A_2	B_2
I	A_1	100 saved	All die
	B_1	All die	All saved

Let us first record the implications of the Simple Principle. It implies that I have one group-based reason *against* performing B_1 and one group-based reason *against* performing A_1 since you and I together have a reason against performing $\langle B_1, A_2 \rangle$ and a reason against performing $\langle A_1, B_2 \rangle$. (For our purposes, we may ignore your reasons.) This is so because each of these group actions would produce suboptimal outcomes if they were performed. According to Dietz, these two conflicting reasons “cancel each other out.”⁶

The Simple Principle also implies that I have a group-based reason in *favor* of performing B_1 since you and I together have a reason to perform group action $\langle B_1, B_2 \rangle$, which we have because $\langle B_1, B_2 \rangle$ would produce the optimal outcome in the circumstances if it were performed. Furthermore, it implies that I have a group-based reason *against* performing A_1 since you and I have a reason against performing group action $\langle A_1, A_2 \rangle$, which we have since $\langle A_1, A_2 \rangle$ would produce a suboptimal outcome if it were performed. These two reasons are not in conflict but rather point in the same direction. Although it will not be important here, it is plausible to assume that these two reasons do not add up.

In view of these remarks, it seems fair to say that the Simple Principle implies that I have one “effective” group-based reason in the case before us, namely, one in favor of performing B_1 .

Now, suppose that you will actually perform A_2 . In that case, apart from having an “effective” group-based reason in favor of performing B_1 , I also have an ordinary reason to perform A_1 since otherwise one hundred more people will die. So, if you perform A_2 , there is a conflict between my group-based reason to perform B_1 and my ordinary reason to perform A_1 .

Since there is this conflict between these two reasons and since, intuitively, I ought to perform A_1 (since otherwise one hundred more people will die), proponents of the Simple Principle need a plausible view on the strength of group-based reasons that generates this result. According to Dietz, it seems plausible that the strength of a person’s group-based reason to perform an action is some proportion of the strength of the group’s reason to perform the group action of which this action is a part.

6 Dietz, “What We Together Ought to Do,” 979.

According to this view, the strength of my group-based reason to perform B_1 is some proportion of the strength of the reason you and I together have to perform $\langle B_1, B_2 \rangle$. But what is the strength of this reason? According to Dietz, it “seems plausible that, if we are in a position to save lives, either together or individually, the strength of our reasons to do so will be proportional to the number of lives that would be saved.”⁷ Consequently, the strength of our reason to perform $\langle B_1, B_2 \rangle$ is one million.

For the sake of argument, Dietz assumes that a person’s group-based reason is “one thousandth as strong” as the group’s reason. Given this assumption, it turns out that I ought to perform B_1 even if you perform A_2 since I then will have a group-based reason of strength one thousand to perform B_1 but only have an ordinary reason of strength one hundred to perform A_1 . I have an ordinary reason of strength one hundred to perform A_1 if you perform A_2 , since one hundred more people will be saved if I under those circumstances perform A_1 . However, since intuitively I ought to perform A_1 if you perform A_2 , the assumption that a person’s group-based reason is “one thousandth as strong” as the group’s reason must be wrong.

However, it would be no remedy to assume that a person’s group-based reason might be a lot weaker than one thousandth as strong as the group’s reason since a version of Impending Disaster may be constructed where many more than one million lives are at risk.

According to Dietz, the view that the strength of a person’s group-based reason is some proportion of the strength of the group’s reason should therefore be rejected. He briefly considers the view that there may be an upper bound on the strength of a person’s group-based reason, but he rejects this view too. He argues that it is *ad hoc*, and that “setting the bound at any particular strength seems arbitrary.”⁸

Having considered this argument, Dietz concludes that the Simple Principle should be rejected.

3. DIETZ’S PRINCIPLE

As we just saw, the Simple Principle stumbles when someone fails to do her part of the group action that would produce the optimal outcome in the circumstances. To avoid this problem, Dietz adopts the following principle instead:

Dietz’s Principle: If a person is a member of a group such that the group

7 Dietz, “What We Together Ought to Do,” 971.

8 Dietz, “What We Together Ought to Do,” 972.

has a reason (not) to perform a group action ϕ , and the group will perform ϕ if the person does her part of ϕ , then that person has a reason (not) to do her part of ϕ .⁹

Dietz's Principle does not face the same difficulty as the Simple Principle faces. Consider again Impending Disaster. First, suppose that you perform A_2 . If so, I have (as noted above) an ordinary reason to perform A_1 since otherwise one hundred more people will die. Of course, I have a group-based reason against performing A_1 , according to Dietz's Principle, since you and I together have a reason against performing $\langle A_1, A_2 \rangle$ and we will perform this action if I perform A_1 (since you perform A_2). However, I also have a group-based reason against performing B_1 since you and I together have a reason against performing $\langle B_1, A_2 \rangle$ and we will perform this action if I perform B_1 (since you perform A_2). And since my group-based reason against doing B_1 presumably is stronger than my group-based reason against doing A_1 , the latter is "cancelled out."¹⁰

Second, suppose that you perform B_2 . In that case, I have an ordinary reason to perform B_1 since a million people will otherwise die. I also have a group-based reason in favor of performing B_1 , according to Dietz's Principle, since you and I together have a reason to perform $\langle B_1, B_2 \rangle$ and we will perform that action if I perform B_1 (since you perform B_2). My group-based reason and my ordinary reason point in the same direction under these circumstances. I also have a group-based reason against performing A_1 since you and I together have a reason against performing $\langle A_1, B_2 \rangle$ and we will perform this action if I perform A_1 (since you perform B_2). This reason is also in line with my ordinary reason to perform B_1 .

4. AN ARGUMENT AGAINST DIETZ'S PRINCIPLE

My counterexample to Dietz's Principle is based on the following case, which I call the "Overdetermination Case": you and I, respectively, have two options. Each of us can either shoot the same person or refrain from doing so. If either of us shoots or if both of us shoot, the person dies, and if neither shoots, the person lives. Furthermore, our actions are counterfactually independent. Our situation is captured by the following:

- 9 Dietz, "What We Together Ought to Do," 977. My four remarks on the Simple Principle in section 1 also apply to Dietz's Principle.
- 10 This is so because the former is derived from a group action that you and I have a stronger reason against performing than the one from which the latter is derived. If there is an upper bound on the strength of a person's group-based reason, they may be equally strong.

	<i>You shoot</i>	<i>You do not shoot</i>
<i>I shoot</i>	The victim dies	The victim dies
<i>I do not shoot</i>	The victim dies	The victim lives

Suppose that you will shoot. In that case, I have a group-based reason against abstaining from shooting, according to Dietz's Principle. First, you and I have a reason against performing <I do not shoot, you shoot> since there is an alternative that has a better outcome. Second, we will perform <I do not shoot, you shoot> if I do not shoot since—I assume—you will shoot. However, intuitively, I do not have a reason against abstaining from shooting.¹¹ This is my counterexample to Dietz's Principle.

Dietz might deny that it is counterintuitive that I have a reason against abstaining from shooting (if you shoot). However, that would not be a very plausible reply. According to Dietz's Principle, I have a group-based reason against abstaining from shooting (if you shoot) partly because you and I have a reason against performing <I do not shoot, you shoot>. Moreover, you and I have a reason against performing this group action because there is an alternative whose outcome would be better. However, my abstaining from shooting does not contribute to the suboptimal outcome (i.e., that the victim dies) associated with <I do not shoot, you shoot>. In fact, it is necessary for the optimal outcome. In light of this, it would be implausible to insist that I have a reason against abstaining from shooting if you shoot.

Dietz might also reply that the fact that I have a group-based reason against abstaining from shooting does not determine what I ought to do and that it may therefore be ignored. This is because I also have a group-based reason against shooting (if you shoot), according to Dietz's Principle. You and I have a reason against performing <I shoot, you shoot> since there is an alternative that has a better outcome. Moreover, we will perform <I shoot, you shoot> if I shoot since—I assume—you will shoot. The upshot is (as above) that these two conflicting group-based reasons "cancel each other out." However, this reply is beside the point. It is still the case that, intuitively, I do not have a reason against abstaining from shooting.

Before I move on, I want to point out (although it might already be evident) that the case discussed in this section may be used against the Simple Principle

¹¹ I think that this is true also if you and I have explicitly agreed to perform <I do not shoot, you shoot> since I nevertheless do not contribute to the victim's death and since your action does not depend on mine.

too. The Simple Principle also implies that I have a reason against abstaining from shooting even if you shoot since you and I have a reason against performing $\langle I \text{ do not shoot, you shoot} \rangle$. Furthermore, I also want to note that Impending Disaster could also be used to raise a similar objection to the Simple Principle. As we saw, the Simple Principle implies that I have a group-based reason against performing B_1 in Impending Disaster (also if you perform B_2) since you and I have a reason against performing $\langle B_1, A_2 \rangle$. However, intuitively, I do not have a reason against performing B_1 (if you perform B_2).

5. REVISING DIETZ'S PRINCIPLE

Dietz might revise his principle. A suggestion that would not be promising is that Dietz should focus on reasons in favor of actions. Dietz might suggest that his principle should only be concerned with reasons in favor of actions and not also (as it is now) with reasons against actions. More precisely, he might suggest that we remove the two occurrences of “(not)” from his principle. I criticized Dietz’s Principle because it implies, counterintuitively, that I have a group-based reason against abstaining from shooting (if you shoot), in the Overdetermination Case. This principle—i.e., the principle you get if you remove the two occurrences of “(not)” from Dietz’s Principle—does not have this implication. As I noted, it is the fact that you and I have a reason against performing $\langle I \text{ do not shoot, you shoot} \rangle$ and that we will perform this action if I abstain from shooting (if you shoot) that together with Dietz’s Principle have this implication. However, as I said in the beginning of this paragraph, this would not be a promising suggestion. It would be an odd asymmetry if people were to inherit reasons in favor of actions but not also reasons against actions. If you inherit reasons in favor of actions it simply seems implausible that you do not also inherit reasons against actions.

However, there is a natural suggestion that might seem more promising. Dietz might suggest that we concentrate on those group actions the person herself can realize given the actual behavior of the other members.¹² Suppose that I can perform action A and action B , and that you can perform action C and action D . Furthermore, suppose that you actually perform C . Under these circumstances, I can realize two group actions, namely, $\langle A, C \rangle$ and $\langle B, C \rangle$. Dietz might then suggest that I have a group-based reason in favor of performing A if the outcome of $\langle A, C \rangle$ is better than the outcome of $\langle B, C \rangle$ and that I have a

12 I am grateful to a referee of this journal for the suggestion that Dietz might try to revise his view along these lines.

group-based reason against performing A if the outcome of $\langle A, C \rangle$ is worse than the outcome of $\langle B, C \rangle$. In short, Dietz might propose the following principle:

The Revised Principle: A person has a reason (not) to do her part of a group action ϕ if that person can realize ϕ given the actual actions of the other members of the group, and the outcome of ϕ is (worse) better than the outcome of any other group action that the person can realize given the actual actions of the other members of the group.

Again, I criticized Dietz's Principle because it implies, counterintuitively, that I have a group-based reason against abstaining from shooting (if you shoot) in the Overdetermination Case. Just as the principle I briefly considered in the beginning of this section does not have this implication, the Revised Principle does not have it either. I can realize two group actions if you shoot: $\langle I \text{ do not shoot, you shoot} \rangle$ and $\langle I \text{ shoot, you shoot} \rangle$. And the former does not have a worse outcome than the latter. Consequently, the Revised Principle does not imply that I have a group-based reason against abstaining from shooting if you shoot.

But there is another problem with the Revised Principle. As with the other principles discussed above, it is a principle about group-based reasons. It is supposed to tell us under what circumstances the members of a group inherit reasons to act from the group. A reasonable requirement for a person to inherit a reason to perform an action from the group is surely that this action is a part of a group action that the group has a reason to perform. A part can hardly inherit a feature from the whole that the whole lacks. According to the Revised Principle, however, a person may inherit a reason to perform a certain action although the group does not have a reason to perform the group action of which the action is a part.

Consider again Impending Disaster, for example. Suppose that you perform A_2 . In that case, I have a group-based reason to perform A_1 . Since you perform A_2 , I can realize two group actions, namely, $\langle A_1, A_2 \rangle$ and $\langle B_1, A_2 \rangle$. And the former clearly has a better outcome than the latter. But from what group action is my group-based reason to perform A_1 supposed to be derived? There is one plausible candidate: $\langle A_1, A_2 \rangle$. However, you and I do not have a reason to perform $\langle A_1, A_2 \rangle$ in the circumstances. The only group action you and I have a reason to perform is $\langle B_1, B_2 \rangle$.¹³ So, according to the Revised Principle, a person may in-

13 Although you and I may have a reason to perform $\langle A_1, A_2 \rangle$ rather than $\langle B_1, A_2 \rangle$, this is not relevant to our concerns here. We are concerned with the reasons we have to perform actions full stop. And in the circumstances, we only have a reason to perform $\langle B_1, B_2 \rangle$ full stop.

herit a reason to perform an action although the group lacks a reason to perform the group action of which this action is a part.

To be sure, the Revised Principle implies that I have a group-based reason against performing B_1 if you perform A_2 . Again, I can realize two group actions if you perform A_2 : $\langle A_1, A_2 \rangle$ and $\langle B_1, A_2 \rangle$. And the latter has a worse outcome than the former. So, it may seem that I have a group-based reason to perform A_1 after all, since A_1 is the only alternative I have to B_1 . Furthermore, importantly, my group-based reason against performing B_1 points in the same direction as the reason that you and I have against performing $\langle B_1, A_2 \rangle$.¹⁴ However, the explanation of why you and I have a reason against performing $\langle B_1, A_2 \rangle$ has nothing to do with action A_1 . We have this reason because $\langle B_1, B_2 \rangle$ has a better outcome.

Finally, note that Dietz's Principle does not suffer from the same shortcoming. It does not imply that I have a reason to perform A_1 if you perform A_2 . As I said, according to Dietz's Principle, a person has a reason to do her part of a group action if the group of which the person is a member has a reason to perform the group action and this group will perform the group action if the person does her part. And you and I do not have a reason to perform $\langle A_1, A_2 \rangle$.¹⁵

Uppsala University
magnus.jedenheim-edling@filosofi.uu.se

REFERENCES

- Dietz, Alexander. "What We Together Ought to Do." *Ethics* 126, no. 4 (July 2016): 955–82.
- Jackson, Frank. "Group Morality." *Metaphysics and Morality: Essays in Honour of J. J. C. Smart*, edited by Philip Pettit, Richard Sylvan, and Jean Norman. Oxford: Basil Blackwell, 1987.

- 14 As we just saw, my reason to perform A_1 does not point in the same direction as our reason to perform $\langle A_1, A_2 \rangle$. According to the Revised Principle, I have a reason to perform A_1 , although you and I do not have a reason to perform $\langle A_1, A_2 \rangle$.
- 15 I am grateful to participants in the Higher Seminar in Practical Philosophy at Uppsala University for very helpful comments. I give special thanks to Erik Carlson, Olle Risberg, Pekka Väyrynen, and a referee of this journal whose comments led to substantial improvements. Finally, I am particularly indebted to Jens Johansson for invaluable comments on most parts of this article. My research for this article was funded by a Göransson-Sandviken Research Scholarship awarded to me by Gästrike-Hälsinge Nation, Uppsala University. I am very grateful to them for their support.

Tännsjö, Torbjörn. "The Morality of Collective Actions." *Philosophical Quarterly* 39, no. 155 (April 1989): 221–28.

Woodard, Christopher. *Reasons, Patterns, and Cooperation*. New York: Routledge, 2008.