# Moral Psychology and the Unity of Morality

JAMES G. QUIGLEY
*Florida State University*

Jonathan Haidt's research on moral cognition has revealed that political liberals moralize mostly in terms of Harm and Fairness, whereas conservatives moralize in terms of those plus loyalty to Ingroup, respect for Authority, and Purity (or IAP). Some have concluded that the norms of morality encompass a wide variety of subject matters with no deep unity. To the contrary, I argue that the conservative position is partially debunked by its own lights. IAP norms' moral relevance depends on their tendency to promote welfare (especially to prevent harm). I argue that all moral agents, including conservatives, are committed to that claim at least implicitly. I then argue that an evolutionary account of moral cognition partially debunks the view that welfare-irrelevant IAP norms have moral force. Haidt's own normative commitments are harmonized by this view: IAP norms are more important than liberals often realize, yet morality is at bottom all about promoting welfare.

Jonathan Haidt has recently defended a broad and influential account of moral cognition called Moral Foundations Theory (MFT).[1] The Theory proposes that the human mind comes specially disposed to moralize in accordance with at least six motifs; these are labelled 'foundations'. As children develop, their particular culture may emphasize some of the foundations more than others, leading to societies with distinct patterns of moralizing.

The major empirical finding is that people across the 'liberal-conservative' political spectrum tend to embody two distinct patterns in their moral judgements.[2] Liberals' moral judgements, in all cultures surveyed, are confined to the motifs of Harm, Fairness and (to a lesser extent) Liberty, whereas those of conservatives additionally take into account the other foundations: loyalty to Ingroup, deference to

[1] Jonathan Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (New York, 2012).

[2] In Haidt's surveys, 'very conservative' and 'very liberal' are usually poles on a linear spectrum of preliminary self-identification. I suggest the labels are best understood as names for two general propensities in moralizing, rather than as political categories. These categories – along with the labels 'Ingroup', 'Authority' and 'Purity' – are vague and probably multifaceted. For example, different kinds of conservatives might moralize about sexual and ceremonial impurities, respectively. However, here there is no need to disambiguate further than Haidt does.

© Cambridge University Press 2015
*Utilitas* Vol. 27, No. 2, June 2015
doi:10.1017/S0953820814000351

CrossMark
https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

Authority and Purity.[3] (Here I will refer to Ingroup, Authority and Purity issues together as IAP issues.) Let us take some examples from the Moral Foundations Questionnaire developed by Haidt. When deciding 'whether something is right or wrong', conservatives but not liberals tend to treat the following as highly relevant: whether or not someone showed a lack of loyalty or betrayed one's group (Ingroup); whether or not someone showed a lack of respect for authority or rather conformed to the traditions of society (Authority); and whether or not someone violated standards of purity and decency or rather acted in a way God would approve of (Purity).[4] And to recount one oft-used vignette: conservatives tend to judge it wrong for a family to eat their deceased pet dog's carcass, whereas liberals tend to say that this not wrong but merely disgusting or disrespectful.[5] Here let us take for granted these findings and the MFT interpretation of them.[6]

Officially, MFT merely describes the psychological phenomena just mentioned. But unofficially the findings tempt MFT theorists and others towards normative claims such as that IAP norms should be followed for their own sakes, or that people should adopt a more politically 'conservative' way of thinking about morality.[7] (I aim to resist these.) For instance, consider two writers' anti-objectivist conclusions. Walter Sinnott-Armstrong uses Haidt's findings to argue that the content of morality is fundamentally disunified: that certain prohibitions widely considered moral – ratting on friends (Ingroup), disrespecting one's parents (Authority) and having consensual premarital sex (Purity) – are moral issues independent of

[3] As confirmed so far in at least eleven cultures. See Jesse Graham, B. A. Nosek, Jonathan Haidt, Ravi Iyer, Spassena Koleva and P. H. Ditto, 'Mapping the Moral Domain', *Journal of Personality and Social Psychology* 101 (2011), pp. 366–85.

[4] The MFQ is used extensively in Graham et al., 'Mapping the Moral Domain', and can be found in that study's appendix as well as at <http://www.moralfoundations.org/sites/default/files/files/MFQ30.self-scorable.doc>.

[5] Jonathan Haidt, Silvia Helena Koller and Maria G. Dias, 'Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?', *Journal of Personality and Social Psychology* 65 (1993), pp. 613–28.

[6] There may be various grounds for debate about the MFT model. Kurt Gray and co-authors take issue, arguing that all moral transgressions, even purity violations as understood by conservatives, are implicitly (subconsciously) understood in terms of agents harming something with a mind. See Gray Liane Young, and Adam Waytz, 'Mind Perception is the Essence of Morality', *Psychological Inquiry* 23 (2012), pp. 104–24. The present article's debunking approach is more resistant to counterexamples and inconvenient empirical results than Gray et al.'s paper. Even if conservatives' IAP judgements consistently depart from Gray et al.'s model, this article contends that conservatives are committed to a Harm-based moral framework as a matter of *logical* implication.

[7] Haidt praises the virtues of conservative moral thinking in places such as *Righteous Mind*, pp. 305–9.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

any connection to harm.[8] Second, Jesse Prinz uses Haidt to argue for a relativist view on which moral truths are indexed to the (often competing) emotional commitments of cultures and individuals. He writes that: 'The same policy can be right on liberal values, but not right on conservative values . . . [and] there is no transcendental position from which one can decide which value system is better.'[9] To the contrary, I argue that MFT is compatible with viewing welfare as a basic moral value that IAP norms promote.

Here are the theses of this article. No IAP considerations are intrinsically morally relevant. Sometimes IAP considerations enjoy moral relevance, but only (and exactly) in so far as they *protect and promote welfare* – primarily by preventing or reducing Harms of one sort or another.[10] In other words:

> *Moral Dependence Thesis (MDT):* Ingroup, Authority and Purity norms apply exactly when, and exactly because, certain norms encouraging welfare promotion (especially norms forbidding harming) apply.

This is to say that IAP norms exhibit what I'll call *normative dependence* on welfare-promoting norms. Furthermore, although some people have the intuition that IAP considerations are intrinsically morally relevant independent of their connection to welfare, this intuition is subject to a debunking argument. (This debunking argument says that IAP intuitions are mostly outdated heuristics for welfare promotion; people who take IAP norms too seriously are blind to this because of the natural human tendency to treat moralized norms as all-important.) Nevertheless, IAP norms have extrinsic moral importance, as they are not completely outdated for our world; in this respect MFT theorists are right in their occasional defences of the conservative moral outlook.

This article could serve as part of an argument that morality is all about welfare promotion (a view sometimes labelled 'welfarism'[11]). But it is not a full argument for welfarism in so far as, for all I say here,

[8] See Sinnott-Armstrong, 'Is Moral Phenomenology Unified?', *Phenomenology and the Cognitive Sciences* 7 (2008), pp. 85–97; and Walter Sinnott-Armstrong and Thalia Wheatley, 'The Disunity of Morality and Why it Matters to Philosophy', *Monist* 95 (2012), pp. 355–77.

[9] Jesse Prinz, *The Emotional Construction of Morals* (Oxford, 2007), p. 206.

[10] In this article I will use 'harm prevention' and 'welfare promotion' as shorthand phrases for the following property of moral rules: their serving to promote welfare, especially by prohibiting and discouraging actions which are harmful or tend to be harmful, and also by recommending or requiring actions which are beneficial or tend to be beneficial.

[11] Such as by Simon Keller, 'Welfarism', *Philosophy Compass* 4 (2009), pp. 82–95.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

other harm-independent subject matters might be intrinsically morally relevant. To be sure, I suspect that Fairness and Liberty considerations (for example) will turn out to exhibit a normative dependence on welfare similarly to how IAP considerations do. But this is not the place to explore such proposals thoroughly.

This article is neutral on at least four contested questions. The first is the debate between consequentialists and deontologists. For all I say, deontologists might be right: there might be special constraints against (e.g.) harming certain persons, or against doing the harming oneself, or there might be duties to benefit persons to whom one is specially related.[12] Second, it is neutral regarding whether (e.g.) lying, promise-keeping or the development of one's talents are morally relevant independent of any connection to the harms or benefits of these practices to anyone. Granted, such views might look more doubtful in light of MDT; but neither am I arguing against such views. Third, this article is compatible with any cognitivist metaethical stance, according to which moral judgements are capable of being true or false, realist or otherwise (more will be said about this in section III). Finally, this article does not engage in debate about whether there is a distinct or unified psychological pattern in virtue of which attitudes count as 'moral' or 'moralistic'.[13] Our topic is about which norms, if any, would generate moral requirements for us (*morality*, in that sense).

Here's a roadmap for the article. First (section I), I will clarify the view. Next (section II), I will support the claim that many other moral norms depend for their moral relevance on their connection to welfare promotion. I will also provide preliminary reasons to hypothesize that virtually all moral agents – including both conservatives and liberals – are at least implicitly committed to that claim. This is some reason to expect IAP norms will also depend on welfare promotion. Third (section III), I will relay a promising evolutionary story of how IAP norms emerged. This story, I argue, vindicates the view that IAP norms were (extrinsically) useful in ancestral, tribal environments. However, it debunks the view that IAP norms have moral force even when they have no connection to welfare protection. Finally (section IV), I'll explain how tracing IAP's moral import to their welfare-relevance accommodates Haidt's apparently disparate normative commitments, and does so better than any relativist view.

[12] For other treatments of welfarism as neutral on this score, see Keller, 'Welfarism', p. 88, and L. W. Sumner, *Welfare, Happiness, and Ethics* (Oxford, 1996), ch. 7.

[13] Debate on that interesting question takes place, for example, in Sinnott-Armstrong, 'Is Moral Phenomenology Unified?'.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

## I. CLARIFYING THE VIEW

### *I.1. Normative dependence*

We often explain the application of certain norms or rules in terms of others. For example, 'the point' of a certain linguistic convention is to disambiguate; that of song-writing advice, to produce aesthetically pleasing songs; that of reasoning advice, to form well-justified beliefs. These are examples of what we can call *normative dependence* – a concept easier illustrated than defined.

Full relations of normative dependence seem to have at least two aspects to them, an asymmetrical explanatory relation and a heuristical relation. Let me illustrate with a chess analogy. The chess stratagem (3) *do not triple your pawns* depends on the stratagem (2) *do not lose material (pieces) relative to your opponent*. (2) is 'the (whole) point' of (3). A chess player committed to not losing the game prevents her pawns from tripling 'for the sake of', '*because* of' (and *only because* of) the importance of this for not falling behind in material. In turn, the whole point of not losing one's pieces is to comply with the twin supreme objects of chess, (1A) *do not allow yourself to be checkmated* and (1B) *checkmate your opponent*.

Similarly, stratagems like (3) are mere rules-of-thumb (heuristics, guidelines): at least approximately, they are to be followed *when and only when* they are conducive to successfully following the more fundamental stratagem.[14] Suppose by tripling your pawns you can win your opponent's rook: then triple away. The rule against tripling one's pawns is to be followed *exactly in so far as* – i.e. because and only because, and when and only when – it conduces to not falling behind in material.

Below it will be argued that IAP norms exhibit a *moral* normative dependence on Harm norms similar to the dependence of (3) on (2). For example, following a rule like (~pms) *do not have premarital sex* is morally advisable when and only when, and because and only because, it is a way of following a more fundamental norm, (~H) *do not harm (or risk producing harm for) others (inexcusably)*. ((~H) is an approximation that will be explicated next.)

[14] There are extremely rare scenarios in which one could triple one's pawns, and lose material, but checkmate one's opponent (or avoid being checkmated). Such cases illustrate the ultimate normative dependence of (3) on (1B). Such cases are the reason the qualifier (3) applies *approximately* when and only when (2) applies.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

### I.2. The foundational norm

Most pithily, the norm (~H) is *do not wrongfully harm others*.[15] That formula is quite unhelpful, but at least it makes two things clearer. First, harmed states in others are things to be prevented or not caused. Second, *wrongful* harming is really what morality proscribes.[16] Many further provisos can explicate wrongful harming. (Notably: a wrongful *act* of harming, *or allowance of harm* must be *intentional or negligent*, with respect to the *risks* of harms that the agent could *reasonably expect* to produce, where the act is *the cause* of the harmed state and is *neither excusable nor justifiable* by considerations such as desert or consent.)[17] The first, simpler point is more important here: causing harmed states in others creates some need for justification.

The parties which moral rules protect from harms may be either determinate individuals or merely someone or other. They could be a group of individuals, or even society in general. Moreover, moral harm prohibitions are presumably broad enough to prohibit inducing some other person, or even some non-agent, to bring it about that someone enters a (more severely) harmed state. (For example, the wrongness of allowing poisonous plants to pollute (imPurify) Pittsburgh's water supply is to be explained by the harmed states it induces.)

### II.3. Harm, benefit and welfare

MDT and welfarism seem compatible with any plausible account of the nature of welfare, harms and benefits.[18] But as we consider (in section III) the kinds of harms on which IAP norms might normatively depend, it will help to have at least a preliminary, fairly neutral account of these notions.[19]

A harmed state is a state in which one's interest is set back.[20] While 'interest' (in this sense) may not be reductively analysable, we can

[15] To simplify, I'll ignore the alternative welfarist thesis that some IAP norms depend on (~H) and other IAP norms depend on a rule like (B) 'benefit others in circumstance x'.

[16] Indeed, we sometimes seem to use a normative sense of 'harm' on which A harms B iff A *wrongfully* puts B in a harmed state. Joel Feinberg, *The Moral Limits of the Criminal Law, Volume 1: Harm to Others* (Oxford, 1984), p. 34.

[17] This list borrows heavily from Feinberg, *Harm to Others*, p. 105.

[18] Cf. Keller, 'Welfarism'.

[19] This account should be friendly both to common sense and to Haidt, who characterizes harm simply by associating it with a cluster of related concepts. These include (physical and emotional) need, suffering, distress; death; cruelty, unkindness; care, compassion; attachment, nurturance, and tender feelings toward cute stimuli. See e.g. *Righteous Mind*, pp. 131–4.

[20] My categories of 'vulnerability-' and 'desire-grounded' interests are parallel to what Feinberg calls 'welfare interests' and 'ulterior interests', respectively (see *Harm to Others*, p. 37). For similar distinctions, see Nicholas Rescher, *Welfare: The Social Issue in Philosophical Perspective*, (Pittsburgh, 1972), and Shlomit Harrosh, 'Identifying Harms', *Bioethics* 26 (2012), pp. 493–8.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

say that having an interest is 'having a stake' in how a thing fares – whether that thing is one's mind, consciousness, leg, mother, romantic relationship, career, hobby, etc. Interests seem to come in two main classes (this will be relevant in section II).

*Vulnerability-grounded interests* are interests in avoiding disturbances to the development and integrity of one's body and mind as a means to the continuance for a foreseeable interval of one's life.[21] As animals we have *basic interests* in virtue of certain basic bodily and mental vulnerabilities. These vulnerabilities (i.e. needs) are to negative experiences (pains, including aspects of some negative emotions) and to disruption of bodily and mental integrity, yielding a list of *basic harms*: death, pain, disability, injury and illness. Other vulnerability-grounded interests are derivative in that they are means to resources required for avoiding basic harms. Universally, our *primitive needs* are for sustaining resources (e.g. oxygen, food and water, sleep, exercise; minimal mental acuity and emotional stability). Similarly, other *common needs* include: shelter; resources for affording sustenance (e.g. hunting implements, money); sanity-preserving sociality (e.g. interpersonal attachment, intimacy, respect); and liberty (e.g. of movement) to pursue means of meeting these other needs.

*Desire-grounded interests* are interests peculiar to individuals which are plausibly grounded in stable and deeply rooted 'desires' (loosely construed so as to include goals, plans, etc.). Intuitively, a desire typically gives rise to an interest only if it is strong/deep, durable, based on a realistic hope or expectation, is a desire for a thing for its own sake, is capable of promotion by the person's own efforts, and is either about oneself or one's self-narrative.[22] (For example, a hope that one's child grows up to be a lawyer could plausibly create a desire-grounded interest, but a momentary wish that my fellow airline passenger find his book does not.) Examples of desire-grounded harms include loss of: reasonably expected benefits, data, vacation time, discretionary income, eligibility for a competition, and networking opportunities.[23]

[21] Feinberg, *Harm to Others*, p. 37.

[22] Feinberg, *Harm to Others*, pp. 44–5.

[23] There are many important questions about the nature of harm which this article cannot and need not settle. A sampling: first, just *to what extent* a desire, to ground an interest, must meet the six conditions we have listed. Second, whether certain entities can be patients of harm – e.g. animals, foetuses, groups of people, organizations, the natural environment, future persons, dead persons, the institution of marriage, etc. Third, whether and how harms are relative to a 'baseline' of normality. Fourth, whether we should distinguish being harmed from being offended (e.g. annoyed, embarrassed) or from other negative emotional or hedonic states (e.g. boredom, foul odours).

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

Benefits also come in two categories. A *need-fulfilling benefit* is simply a good for which a creature has a need (e.g. food, shelter, a resource, a companion). Second, a *pleasure-providing benefit* is an experience of pleasure (including aspects of some positive emotions). An event might provide need-fulfilment and/or pleasure directly (a basic/intrinsic benefit), or indirectly (a derivative/extrinsic benefit). Examples of derivative benefits range from the very general (freedom, ability, opportunity), to the primitive or nearly universal (e.g. consciousness, minimal physical exercise; play, sex, intellectual engagement, novel experience, etc.).

A thing's *welfare* can be construed as an overall measure of how it is faring with respect to (the harmonious advancement of) its constellation of interests. Countless interesting puzzles about welfare cannot be answered here.

### *I.4. Caveat on 'intensifiers'*

Here is a final pair of elucidations before we defend the view. I can allow that IAP considerations could still be morally relevant as intensifiers of the harmfulness of certain acts of harming. Murdering an Ingroup-member, or a (legitimate) Authority, rather than a stranger, might reasonably be construed as a greater harm. On one version, we might suppose that the victim has desires regarding special treatment by the murderer, constituting an interest which is then set back, and this counts as an extra harm beyond the harm constituted by the killing. Alternatively, someone might decide the notion of harm is sufficiently flexible that merely disrespecting these relationships counts as a special form of harm (although I would want an explanation for this).

Now someone might insist that IAP considerations are morally relevant not in those ways, but merely as intensifiers of the *wrongness* of certain acts of harming. ('Reasons are holistic!', they might insist.[24]) Even if this point could be well motivated (and I am sceptical), I think it is *largely* compatible with my position. This point says that IAP considerations do not have moral weight when welfare considerations are 'out of the picture'; they only have moral weight after harm has 'appeared in the picture'. I can maintain that IAP norm violations are wrong *when and only when* they have some (reasonable) connection to welfare-hindrance and that they are wrong *because* they have such a connection. I need only concede that, in a few cases, IAP norm violations are sometimes wrong not *only because* of the harmfulness, but also

[24] The general thought is that IAP considerations might be a reason in one case but no reason at all in another case, or even a countervailing reason. For discussion of holism about reasons see Jonathan Dancy, 'Holism in the Theory of Reasons', *Cogito* 6 (1992), pp. 136–8.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

because (say) of who is getting harmed (e.g. an Ingroup member), or because of how the agent has harmed him (impurely, say), etc. In a few outlying cases, when an act violates both a Harm and an IAP norm, the IAP considerations 'kick in' and (for some reason) add extra normative weight. Requisite modification to my thesis would be as slight. Moreover, the modification is consistent with the main practical upshots of this article's argument, which are (first) that when IAP violations can reasonably be expected not to produce any harms or harmings, they are not wrong, and (second) that where IAP norms are counterproductive to the protection and promotion of welfare, they ought to be abandoned, ignored, or otherwise treated as morally irrelevant.

## II. AGREEMENT ABOUT THE MORAL DEPENDENCE THESIS

Suppose that stereotypical conservatives are for the most part settled on the view that Ingroup, Authority and Purity considerations are fundamentally morally important in their own rights.[25] When presented with IAP violations that are irrelevant to anyone's welfare, suppose, they merely cite the fact that the acts violate IAP norms. ('Anti-patriotism/mockery of elders/zoophilia? Wrong because disloyal/disrespectful/impure! That's all there is to it!') Label this view:

IAPF IAP norms are fundamental explanations of the wrongness of certain acts.

IAPF and MDT cannot both be true. MDT reduces the ultimate grounds of moral rules to welfare promotion, but IAPF denies as much.

There are good reasons to expect that all competent users of moral terms, including conservatives, are committed to at least four theses. These theses, I suggest, logically commit them to MDT, in so far as MDT is the best explanation of these theses. To discover the theses, we should consider ways in which appeals to harm and welfare pervade moral discourse, condemnations and explanations. I will cite available empirical evidence regarding the intuitions of conservatives; however, the relevant empirical literature is so young that these will have to be considered (plausible) hypotheses. But at the least, our purpose is

[25] The basis for this assumption comes from studies which ask conservatives about justifications of moral rules. The best example is Haidt et al., 'Affect, Culture, and Morality'. Haidt expounds this view in *Righteous Mind*, as well as 'The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment', *Psychological Review* 108 (2001), pp. 814–34. This assumption would be worth questioning elsewhere; I am aware of no study in which conservatives are directly asked, 'but why is it bad to do something that violates that kind of Ingroup/Authority/Purity norm?'

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

to show how section III's debunking argument *would* work if it *were* strengthened by the present section's hypotheses. Moreover, even if it turns out that conservatives' commitment to one or two of these theses is limited, this will simply leave a little more work for debunking explanations of why they do not share commitment to the theses – explanations which the debunker may well be able to come by.

Let us consider the four theses and then observe how, together, they support MDT. The first we will call 'HimpM' for the encapsulating phrase 'harm is important to morality':

1. HimpM Actions in which one person harms another (at least, intentionally and without excuse) are (A) an indisputably importantly wrong kind of action, (B) among the most serious moral violations, and (C) wrong even if authorities or (D) cultural customs allow them. (The same is not, or not clearly, true of IAP norms.)

Welfare promotion is (at least part of) the centre of what morality is about. Initial reflection makes sense of why. We humans are all relatively equally vulnerable to setbacks to our interests, and we are deeply concerned with protecting ourselves and those we love from such harms.

If you find HimpM intuitive, you are not alone. Everybody, including conservatives, seems to agree that Harm/welfare (as well as Fairness/justice) matter morally.[26] People of a wide variety of ages, nationalities and religions apparently view norms pertaining to harm, fairness and justice as more (B) serious, (C) authority-independent and (D) culture-independent than violations of merely conventional norms.[27] The evidence is contained in scores of surveys on the so-called 'moral/conventional' distinction, conducted on people of various ages, cultures and nationalities, by Elliot Turiel and several colleagues.[28] By

[26] According to at least two lines of evidence. First, recent internet surveys. In Haidt and colleagues' (Graham et al., 'Mapping the Moral Domain') survey of people across political ideologies in eleven world regions, every political category on average assigned relevance of between 'somewhat' and 'very' to considerations relating to Harm (e.g. cruelty, infliction of suffering) and Fairness (e.g. discrimination, denying people their rights). Second, anthropologists have apparently confirmed that harm prohibitions are culturally universal. See Shaun Nichols, *Sentimental Rules* (Oxford, 2004), p. 142.

[27] The moral/conventional distinction(s) apparently emerge in children as young as 3.5 years. It has been documented in children, adolescents and adults in numerous cultures including Brazil, China, India, Indonesia (preschool children both Muslim and Christian), Israel (both Arab and kibbutz Jewish children), Korea, Ijo children in Nigeria, and Zambia. Autistic children have also been observed to perform normally on the standard moral/conventional distinction. For an overview see Larry Nucci, *Education in the Moral Domain* (Cambridge, 2001).

[28] Knowledgeable readers may now be enticed by recalling recent studies which show that IAP norms elicit some or all of the moral attitude profile ( (A)–(D) ) in certain groups,

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

contrast, IAP norms are much more controversial; liberals are less likely to consider IAP violations to be wrong (although at least in the case of Purity violations they still seem likely to express that the violation *bothers* them).[29]

2. Explain 1 That $\phi$-ing is harmful is an explanation for why $\phi$-ing is morally wrong, for a wide variety of kinds of $\phi$-ing. (This includes varieties of $\phi$-ing which are violations of Ingroup, Authority and Purity norms.)

Declare that there is a 'moral' rule against a certain activity, and people will want to know why, and why it is important. One of the most compelling, popular answers – even for conservatives – will surely be the following. *If everyone felt free to break this rule, many harms, rights violations, or missed or lost benefits would result*. This answer works as an explanation for a wide variety of rules people consider moral: rules against specific forms of harmful behaviour (e.g. hitting or insulting others); trust-maintaining rules (e.g. against deception, theft and cheating); rules of coordination (e.g. traffic regulations); etc. It also seems to explain many IAP norms – for example, it provides an informative, satisfying answer when someone asks why she should 'be loyal to (her) ingroup', 'respect authorities', or should 'not contaminate the atmosphere'.

I can adduce three lines of evidence that conservatives pervasively use harm or welfare to explain what is wrong with many kinds of wrongdoings, including IAP norm violations. First, welfare-citing justifications of IAP norms are interestingly common in public discourse, even by conservatives. For example, regarding Purity norms, one commonly hears appeals to the harm in various sexually and socially deviant practices (miscegenation, masturbation, pornography, gay marriage, etc.).[30] But the evidence is not all anecdotal.

especially certain conservatives, non-westerners, and others. That is simply a distraction concerning the periphery of the category of norms that elicit 'moral' attitudes. There is no question about the point being made in the main text – that harm norms are in within the core of that category. The most noteworthy recent criticism of the moral/conventional distinction is Daniel Kelly, Stephan Stich, K. J. Haley, S. J. Eng and D. M. T. Fessler, 'Harm, Affect, and the Moral/Conventional Distinction', *Mind & Language* 22 (2007), pp. 117–31. For an apt response, see Alejandro Rosas, 'Mistakes to avoid in Attacking the Moral/Conventional Distinction', *The Baltic International Yearbook of Cognition, Logic and Communication* 7 (2012), pp. 1–10.

[29] Regarding the first claim, see Graham et al., 'Mapping the Moral Domain'; regarding the parenthetical claim, see Haidt et al., 'Affect, Culture, and Morality', and cf. Haidt, *Righteous Mind*, p. 95.

[30] For example, see Family Research Council, 'The Top Ten Harms of Same-Sex Marriage', pamphlet available at <http://downloads.frc.org/EF/EF11B30.pdf>. Other anecdotes of appeal to harm, and some indirectly suggestive empirical studies, are provided by Gray et al., 'Mind Perception is the Essence of Morality', p. 108.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

Second, in Haidt's systematic interviews of subjects presented with anecdotes involving harmless IAP violations, subjects have often invoked harm even for cases where the acts have virtually no potential impact on welfare.[31] Regarding his first major study, Haidt avers:

> The biggest surprise was that so many subjects tried to invent victims. I had written the stories carefully to remove all conceivable harm to other people, yet in 38 percent of the 1,620 times that people heard a harmless-offensive story, they claimed that somebody was harmed. In the dog story, for example, [in which a family decides to eat the meat of their recently-deceased family dog,] many people said that the family itself would be harmed because they would get sick from eating dog meat.[32]

Third, various studies have uncovered a phenomenon known as 'moral dumbfounding'.[33] This is where a person (i) judges that an action is wrong, (ii) feels some need to explain why (e.g. because someone else has solicited an explanation from her), but (iii) she cannot find an explanation (or at least is not articulate enough to provide one), yet (iv) she maintains her judgement anyway. The fact that subjects tend to become dumbfounded is itself evidence that they agree with Explain 1. They seem to be sensing that their moral judgement needs some justification, and that appeal to mere feeling or to IAP will not do the trick, but that appeal to welfare would, if only it could be appealed to for the case in question.

Here someone might worry that this results from some bias, due to pressure either from subjects' interviewers, or from their society's general moral ethos, to appeal to welfare in justifying moral claims. (The worry might continue: suppose subjects had been presented with harmful-but-not-offensive acts, whereupon the interviewer would demand justifications in terms of impurity or betrayal.) However, there are two reasons to predict that welfare would be privileged as an important moral explanation even in a variety of interview environments. First, some evidence suggests that if a typical person invokes welfare as justification, she is then insulated against uncertainty when bullied to justify her judgement, and even becomes

[31] Daniel Jacobson presents plausible arguments that the actions probed in Haidt's 2000 'dumbfounding' study are in fact all morally wrong because they are potentially harmful, although not obviously so. Jacobson, 'Moral Dumbfounding and Moral Stupefaction', *Oxford Studies in Normative Ethics*, vol. 2 (Oxford, 2012), pp. 289–316.

[32] Haidt, *Righteous Mind*, p. 24.

[33] The relevant studies include: Jonathan Haidt and Matthew A Hersh, 'Sexual Morality: The Cultures and Emotions of Conservatives and Liberals', *Journal of Applied Social Psychology* 31 (2001), pp. 191–221; and the widely cited but unpublished Jonathan Haidt, Fredrik Björklund and Scott Murphy, 'Moral Dumbfounding: When Intuition Finds No Reason', unpublished manuscript, available at <http://www.faculty.virginia.edu/haidtlab/articles/manuscripts/haidt.bjorklund.working-paper.when%20intuition%20finds%20no%20reason.pub603.doc>.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

more certain about the moral correctness of her judgement in virtue of her appeal to welfare. (Specifically, for one probe in Haidt's 'Moral Dumbfounding' study – the traditional story of Heinz, who steals some medicine needed save his wife's life – subjects became *more* firmly committed to their original judgement that Heinz's action was morally correct, whereas they reversed their judgements on the other cases in the study, which were (putatively) welfare-irrelevant IAP violations.[34]) Second, the truth of HimpM is a simple, general reason to think it would be very surprising to find a culture where harmfulness does not often figure in the explanations of condemnations of a wide variety of acts (even if the concept invoked is broader, e.g. 'sin').

For that matter no one has tested (to my knowledge) the plausible hypothesis that subjects would agree with *legitimately well-argued* claims that certain IAP norm violations do have some real but non-obvious connection to harm prevention or welfare promotion.[35] It may well turn out that if, for example, we ask a conservative, 'Why is it important to love your country?', then, if she does not already appeal to the welfare of the country's constituents, she will at least agree if that answer is suggested to her. At least, it would be very surprising if she insisted, 'No, no, the welfare of the constituents has nothing at all to do with it!'

Notice that, by contrast, IAP norms do *not* seem to work as justification for Harm norms. That is:

3. Explain 2 Explanations of IAP norms' force in terms of welfare promotion usually are satisfying/regress-stopping, substantive, theory-unifying explanations. (However, explanations of welfare-promoting norms in terms of IAP norms usually are *not* satisfying/regress-stopping, substantive, or unifying explanations.)

*Regress-stopping.* Some possible answers to the question, 'Why is it wrong to $\phi$?' don't sufficiently answer the question. '$\phi$-ing tends to produce glacial melting' will be met with, 'So what?', because competent moral thinkers aren't committed to the moral significance of glacial melting. Unless they are, a regress emerges ('. . . yes, but so what?') potentially *ad infinitum*.

[34] That is, the interviewers, trained to play 'devil's advocate', 'did change some people's minds, in the direction for which he was playing devil's advocate, except that *on the Heinz story the percentage endorsing Heinz' theft rose even though the interviewer was in most cases arguing against that position.* The percentage of participants who changed their minds averaged 16%, and did not differ significantly across tasks' (Haidt et al., 'Moral Dumbfounding', p. 11, emphasis mine).

[35] Such as the arguments provided in Jacobson, 'Moral Dumbfounding and Moral Stupefaction'.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

The answer 'It is wrong to $\phi$ because it will seriously harm someone' seems always sufficient to stop the regress. Someone who fails to see how harm counts as a reason against $\phi$-ing seems impaired in his comprehension of the terms used, or of moral concepts, or fails to apprehend moral reasons. By contrast, suppose the 'why?' question is answered with, '. . . because $\phi$-ing is disloyal to your ingroup', '. . . is disrespectful to your authorities', or '. . . is impure'. It would display no lack of comprehension to then insist, 'okay, but so what; what moral importance does that have?' Welfare thus bottoms out moral explanations in a way that IAP do not.

*Substantive.* Explaining the wrongness of an IAP norm violation in terms of welfare protection is clearly a substantive explanation, providing new information about the reasons behind certain prohibitions and requirements. We really learn something about moral reasons, for example, when we are told that releasing certain fumes can contribute to future harm and suffering due to global warming.

One might think that the reverse kind of explanation could be substantive. For there are cultural viewpoints – and philosophical theories – which attempt to base moral explanations on some IAP-type property, such as sin, desecration, disobedience to the gods, idolatry, tapu or natural law. You shouldn't harm others, '. . . because it reflects poorly on your ingroup', '. . . goes against an authoritative command', or '. . . violates the sacred order'. But these tell us less about moral reasons than it may initially seem: these views simply seem to get the explanatory primacy wrong. It can sensibly be asked why *those* harmful activities are the ones prohibited by the gods (or by natural law, by my group's honour code, etc.). Welfare threatens to be invariably at the core of the explanation. Circularity is the obvious problem: this view must say these harmful acts are wrong because my group's honour code (e.g.) forbids them, and this code forbids them because they are harmful. The appeal to the putatively basic IAP feature may turn out to be explanatorily redundant; appealing to it would be superfluous. A second potential problem is that some such views seem to lack an informative account of the putative basic notion – e.g. of 'sin' or 'the natural order' – that is really any thicker or more substantive than mere wrongness; 'because it is sinful' comes to little more than 'because it is just wrong'. Third, these accounts may well take nothing away from the welfarist picture in that all the actions prohibited by a welfarist theory will also end up being prohibited in the same ways and degrees. (Or if they do, they do so on pain of implausibility.) From there, if an IAP-based view adds no further moral rules, its rules are simply extensionally equivalent to welfarism. More often, IAP-based views do add further norms, facing the final two worries. The fourth worry is that these welfare-irrelevant moral rules will depend on highly

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

arbitrary factors – the contingent whims of the gods or the tribal leaders – but that neither moral norms in general, nor harm-based norms in particular, seem to be arbitrary in this way. The fifth problem is that IAP-based, welfare-irrelevant norms tend, in practice, to be based on specious claims. (Take, for instance, the claim that cows are holier than pigs, and that widows must not eat fish.[36]) Perhaps not all such claims are vague, unsupported or based on pseudoscience – but many are. Altogether, IAP-based views are prima facie unpromising.

*Theory-unifying*. The unity of morality on the welfarist picture suggested by MDT is a virtue in at least two ways. First, it helps tidy up an account of morality's emergence over the course of human evolution and history, since presumably a system of rules with a single objective emerges more easily than one with multiple objectives. Second, it helps resolve conflicts between subsidiary rules (at least in cases where one rule is clearly more welfare-promoting; more on this below). By contrast, accounts of morality based on more than one IAP norm lack this virtuous unity. (Accounts on which morality has some single point that fits under an I, A or P heading meet the objections noted in the previous paragraph.)

The potential worry about all this tidiness, of course, is that it will fail to include some subtle details about what morality is about. However, the suggestion that IAP rules serve as heuristics for welfare promotion promises to allay some of these worries. It, in turn, is supported by:

4. Override In dilemmas where someone can either Harm a moral patient or violate an IAP norm, it is usually better to violate the IAP norm.

The heuristical aspect of normative dependence has it that the more fundamental norm trumps the dependent norm in cases of conflict. For example, if (~ehf) do not eat human flesh depends on (~H) don't inexcusably cause others to be harmed, then in cases when I would harm someone by eating human flesh, I should not eat such flesh.[37] One survey has found a systematic pattern of intuitions consistent with Override, in both liberals and conservatives. Subjects were shown pairs of descriptions of norm violations of types Harm, Fairness, Ingroup, Authority and Purity – e.g. 'Treating people unequally' (Fairness)

[36] These are examples of IAP norms from Orissa, India, taken from Richard Shweder, Manamohan Mahapatra and Joan G. Miller, 'Culture and Moral Development', *The Emergence of Morality in Young Children*, ed. J. Kagan and S. Lamb (Chicago, 1987), pp. 1–83.

[37] And, complementarily, in the (rare) circumstance in which I will surely in no way harm anyone by violating the rule (~ehf), then I have no reason to follow that rule – just as I have no reason not to triple my pawns when it surely won't lead to loss of material or of my chess game.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

versus 'Disobeying an authority' (Ingroup). Whether instructed to answer carefully and deliberately, or merely as quickly as possible, results indicated that:

> [P]articipants across the political spectrum tend to rank-order foundations in the order HFIAP; strongest preferences were for Harm over Authority and Purity, weakest was for Authority over Purity. In only one case, Fairness vs. Ingroup, did conservatives show a preference opposite to that of liberals (with moderates showing no preference).[38]

If HimpM, Explain 1 and 2, and Override were true, they would together provide a good explanation for MDT, as I will now explain. By contrast, the rival view, of morality as having multiple, irreducible subject matters, accommodates IAPF at the expense of ignoring these four premises. Consider in turn MDT's four parts: IAP norms apply *because*, *only because*, *when*, and *only when*, they are conducive to welfare promotion (especially harm prevention).

The conclusion that IAP norms apply *because* they are welfare-promoting is identical to the latter half of Explain 1. That all IAP norms apply *only because* they are welfare-promoting is well explained especially by Explain 2, as well as by Override and HimpM. IAP norms don't stop explanatory regresses but Harm norms do (Explain 2), so we should conclude that IAP norms can only be fully explained by appeal to welfare-promoting norms. That Harm norms take precedence over IAP norms (Override) suggests that there will not be cases in which IAP norms override Harm norms and thus need some welfare-independent explanation. Finally, the transcultural seriousness of welfare-promoting norms (HimpM) explains why Harm norms stop the explanatory regresses (and more on this below). Once we admit all this, it seems suspicious to hold that IAP norms are independent points of morality.

It is easy enough to show that IAP norms apply *when* Harm norms do: it is as easy as imagining actions which violate both IAP and Harm norms (e.g. torturing your boss). We also might support the thesis that IAP norms apply *especially when* Harm norms do by appeal to the relevant intuitions of conservatives.[39] One place to look would be to the data supporting Explain 1 – i.e. to cases where it seems reasonable to ask why IAP norms should not be violated, especially ones where

[38] Jesse Graham, 'Left Gut, Right Gut: Ideology and Automatic Moral Reactions' (PhD thesis, University of Virginia, 2010). Graham does note that, as expected, liberals were seen to have stronger preferences than moderates and conservatives for Harm and Fairness over Ingroup, Authority, and Purity.

[39] Of course, the most direct support for the 'especially when' thesis would come from conducting an experiment directly examining the additive effects on conservatives' intuitions of combining IAP- with Harm-norm violations. For example, one could ask subjects about (1) torturing a stranger out of contempt versus (2) insulting your father out of contempt versus (3) torturing your father out of contempt – looking for an effect beyond the mere additive effect of (1) + (2).

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

'because it is harmful' is a satisfying answer. These are cases where the explanatory factor adds something deeper or more significant to the wrongness. Override is also suggestive on this front, as it suggests that even conservatives view IAP norms as applying *except when* they are overridden by Harm norms. If we can infer from this that if harmfulness adds something significant to the wrongness of a violation, we could then expect that when an action is both Harmful and an IAP violation, harmfulness as an explanatory factor would also do work in that sort of case, i.e. 'especially when' IAP violations are harmful.

The thesis that IAP norms apply *only when* Harm norms do seems to be the best explanation of all previous points in this section, taken together. If IAP norms apply only because of their conduciveness to welfare promotion, then there is no reason to comply with IAP norms in cases where they in no way conduce to welfare. Conservatives' ambivalence and dumbfounding about welfare-irrelevant IAP violations is also revealing. Their guts tell them something is disturbing about the cases with which they are being presented, and they are motivated to justify the gnawing feeling. But they also surely understand how welfare considerations are a prevalent, obvious, probably primary currency of moral explanations, and the stutters and pauses characteristic of dumbfounding suggest they are least tempted to dismiss the IAP-supporting intuition.

What if we come across a case in which conservatives judge in accordance with an IAP norm but do not feel the need to explain the judgement in welfare-relevant terms, even when pressured to do so? Once we ensure that the subjects do not have some relevant, unobvious belief about harms involved, we must conclude that their righteous minds have duped them into assigning excessive importance to a norm that is peripheral to morality. That is not to say the norm need be *un*important; we can admit that it is in many ways enriching to follow IAP norms, even if they are not-quite-as-important-as-real-morality. There may also be various cases where they happen to encourage moral behaviour. The next section explains how this only-somewhat-dismissive attitude can be justified.

## III. DEBUNKING IAP INTUITIONS

### *III.1. What is debunking?*

Debunking arguments seek to show that someone's apparent justification for his belief is undermined by some considerations about how that belief was formed.[40] Formally, they look like this:

[40] Evolutionary debunking arguments of moral beliefs have been very popular recently. A good overview is Guy Kahane, 'Evolutionary Debunking Arguments', *Noûs* 45 (2011), pp. 103–25. I borrow his formula in what follows.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

| | |
|---|---|
| *Causal premise* | The appearance to S that x is F is (causally) explained by X. |
| *Epistemic premise* | X is an unreliable process. |

Regarding the present topic, I will argue:

| | |
|---|---|
| *Causal premise* | Conservatives' intuitions that welfare-irrelevant IAP norm violations are morally wrong are driven by (i) the fact that IAP norms are ingrained in our cultural histories (and also probably in our bio-cognitive setups) as heuristics for reducing harm or promoting welfare, and by (ii) the fact that humans tend to maintain their original judgements even in the face of countervailing or clearly undermining defeaters, especially on questions about which they have attitudes of moralistic force. |
| *Epistemic premise* | Said intuitions are unreliably formed in virtue of fact (i), because IAP norms are fairly inaccurate heuristics for welfare promotion in contemporary environments, and in virtue of fact (ii), because stubbornness about a belief tends to blind one to defeaters of that belief. |

Let us elaborate, starting with a discussion of (i) from each premise, then of (ii).

### *III.2. (i) in the causal premise: the evolution of IAP norms*

It is plausible that IAP norms emerged as heuristics for promoting welfare, on any plausible account of the evolution of human cultures. First let us consider Ingroup and Authority (IA) norms – giving two reasons – and then move to Purity norms.

A first reason to think that IA norms are heuristics for welfare promotion is simply that the attitudes they require and recommend just are certain varieties of respect for certain parties' welfare. Loyalty to Ingroup is *constituted* by the adoption of attitudes that display a willingness to promote the group's welfare, even at the cost of a substantial sacrifice to oneself.[41] Similarly, respect for Authority is constituted by the adoption of attitudes that display a willingness to protect and promote the welfare of a superior, even at a cost to the welfare of the authority's underlings.

[41] It is a side issue whether the welfare of the group must be understood in terms of the welfare of the individuals who comprise it. Either way, welfare is being promoted.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

A second reason to think that IA norms are heuristics for welfare promotion is that in our evolutionary history they plausibly tended to *cause* people, organized into tribal groups, to be better off than tribal groups with less stringent IA norms. Haidt tells an evolutionary story of inter-tribe hostility, which invokes cultural group selection:

> Humanity's ancestors have been living in groups with at least occasional violent intergroup hostility for most or all of the last seven million years . . . Human beings therefore can be expected to have many ancient 'inside the head' mechanisms (such as for coalitions, tribalism, and territoriality . . . that co-evolved in more recent times with 'outside the head' cultural creations (such as law, religion, and political institutions) to serve the function of suppressing selfishness and increasing group cohesion, trust, and coordinated action.[42]

Group cohesion, in an environment of tribal warfare, is important precisely because it protects group members from harms inflicted by enemies, and allows them to secure benefits for the tribe. Coordinated action is enhanced by competent executive decision, which is in turn fostered by a tendency for group members to defer to a competent authority. Haidt's account therefore implies that IA norms function to promote the welfare of tribal ingroups.[43]

Unless we appeal to IA norms as a cultural adaptation, it is unclear what would explain humans' remarkable ultra-sociality relative to our primate cousins. Within just a few thousand years, Haidt explains,

> humans began engaging in large-scale cooperation, with non-kin, on 'difficult joint projects such as hunting buffalo, weaving large fishnets, and defending

[42] Jonathan Haidt and Selin Kesebir, 'Morality', *Handbook of Social Psychology*, ed. S. T. Fiske, D. T. Gilbert and Gardner Lindzey, 5th edn. (Hoboken, NJ, 2010), pp. 797–832. Two citations have been removed from this quotation.

[43] The point is plausible if Haidt's evolutionary picture is resisted in a couple of ways. First, some readers might resist the appeal to group selection, saying that IA traits really facilitate individual reproductive success rather than group reproductive success (i.e. population growth) or cultural dominance. However, even if so, individuals would achieve reproductive success by avoiding harms and securing resources and other benefits before and during their reproductive stages of life. So IA norms still turn out to be heuristics for welfare promotion. The debate over group selection, or over the thoroughness or sincerity of humans' disposition to sacrifice for their groups, turns out to be a distraction from this article's thesis. (Moreover, Haidt and Kesebir argue persuasively that 'there is now a widespread consensus that cultural group selection occurs' ('Morality', p. 818).)

Second, someone might remind us that welfare protection is not itself the 'goal' of evolution; rather, it is reproductive success (in the case of genetic evolution; and it is something like cultural influence in the case of cultural evolution). However, the protection at least of basic welfare is virtually a necessary condition for reproductive success (as well as for cultural influence). Organisms are prolific to the extent that, before and during their reproductive phases, they are spared death, injury, debilitating pain, etc. and secure resources and other benefits for themselves and their offspring (this is true both for reproductive success and for cultural influence). Furthermore, IA norms just do have respect for and promotion of the welfare of groups and/or their constituents as their explicit goals (to the best of my ethnographic knowledge), with sexual or cultural fertility being construed merely as one important aspect of the group's welfare.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

territory. And once humans domesticated plants and animals and began living in larger and denser groups, they began to engage in large-scale cooperative projects such as building city walls, changing the course of rivers, and conquering their neighbors. . . . The enormous and accelerating gains from cooperation in agriculture, trade, infrastructure, and governance are an example of what has been called a 'major transition' in evolution, during which human beings went from being a social species like chimpanzees to being an 'ultrasocial' species, like bees and ants, able to live in groups of thousands with substantial division of labor.[44]

IA norms (and cognitive dispositions) played (and play) the important role of suppressing selfishness by making people loyal (at least superficially) to their tribal organization and the authorities which organize it. The appeal to the advantage of IA norms in a scenario of inter-tribal competition provides an explanation for why we have IA norms and supporting emotions that are more pervasive than those of our primate cousins.[45]

Purity norms were also plausibly heuristics for welfare promotion. Early humans gradually evolved an ever-elaborate mechanism of disgust, especially as they began to eat meat some 1.5 million years ago.[46] As Daniel Kelly has argued, the disgust response was later plausibly co-opted to play sundry motivational roles, including reinforcing norms relating to ceremonial, spiritual, sexual and ethnic purity (e.g. sexual taboos against incest, intermarriage, sacrilege and ceremonial impurity).[47] Here again, however, the evolutionary contribution of disgust and of purity norms is the protection of the welfare of the bodies and minds of individuals.

*III.3. (i) in the epistemic premise: why IAP get (partially) debunked*

IAP norms in modern environments are fairly inaccurate – but not totally useless – heuristics for impartial welfare promotion. First, consider how IAP heuristics are less accurate now than in ancestral environments. In many ancestral environments, mutual trust and allegiance among an Ingroup would have been the only recourse against crime and treachery; submission to wise Authority the only feasible

[44] Haidt and Kesebir, 'Morality', p. 809. Numerous citations have been removed from this quotation.

[45] As Haidt and Kesebir explain ('Morality', p. 810), 'there is no evidence that any non-human animal feels shame or guilt about violating such norms – only fear of punishment . . . Humans, in contrast, live in a far denser web of norms, mores, and folkways . . . and have an expanded suite of emotions related to violations, whether committed by others (e.g. anger, contempt, and disgust) or by the self (e.g. shame, embarrassment, and guilt)'.

[46] An excellent account of the evolution of disgust is Daniel Kelly, *Yuck! The Nature and Moral Significance of Disgust* (Cambridge, Mass., 2011), ch. 2. See also Paul Rozin, Jonathan Haidt and C. R. McCauley, 'Disgust', *Handbook of Emotions*, ed. M. Lewis, J. M. Haviland-Jones and L. F. Barrett, 3rd edn. (New York, 2008), pp. 757–76.

[47] See Kelly, *Yuck!*.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

policy for eliminating faction or braving the elements. Ancestral people also lacked scientific understandings of, and technologies for combating, diseases and curses. Elaborate purity norms were often the most useful available recourse; better safe than sorry.

Once societies became very large and public, IAP norms lost much relevance. Think of contemporary Western nation-states. As for Ingroup loyalty: most segments of modern industrial and post-industrial societies have little need for pervasive alliances and group loyalties. The outdoors, war, faction, and even crime are less threatening. Instead, hardened group mentalities often encourage competitive, discriminatory, supremacist, jingoistic mentalities – dangerous sentiments in a world with high-powered weapons. As for Authority: where stable democracy is possible, there happens to be less need for rigid, involuntary or non-meritocratic hierarchy as a means of preventing insubordination. Thus, there is less need for the kinds of harmful oppression that often result from giving too much power to a small set of persons. As for Purity: science now gives us much more accurate information about the spread of real contagions and many means of preventing them. There is no longer need or support for beliefs in literal impurities involved in behaviours such as sex and improper ritual.

However, IAP norms are not useless. They can still be useful rules of thumb for us today in places like battlefields, offices and high schools: much harm is reduced through loyalty to friends, respect for teachers, and mitigation of promiscuous or adulterous sexual experimentation. (We will finish this thought in section IV.)

To show how this iconoclasm is justified, I will now answer three key questions, hoping to show that welfare promotion's moral importance is intrinsic, but that that of IAP norms is merely extrinsic.

### *III.4 Questions about the normative force of Harm and IAP*

*1. What's the big deal about welfare?* Section II tried to show that if anything matters morally, it is welfare (the welfare of whoever it is who turns out to matter). This is a priori plausible, and even conservatives seem to agree about this in various ways. But here it is very natural to ask, 'why is welfare important?' This question is ambiguous between at least two questions.

A first version of the question is, 'why should we think anything at all, including welfare, matters morally?' We must delegate this question to articles in which there is room to defend moral realism.[48] Here we are

[48] Recent important exchanges on moral nihilism have featured Richard Joyce and Sharon Street on one hand, and various detractors on the other. For example, see Richard Joyce, *The Evolution of Morality* (Cambridge, Mass., 2006) and Sharon Street, 'A

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

only concerned with discerning what *would* matter morally if anything did. Even if morality turns out to be a fiction or a construction, the contention is that human nature forces us to adopt a standpoint on which welfare-promoting norms are central to morality whereas IAP norms are merely derivative.

A second version of the question is, 'is it a brute fact that welfare matters morally, or can this fact be helpfully explained?' A promising account can be roughly sketched here, as a means of showing how it might turn out; but please note that this article's argument does not depend on its details. Basic harms count as such because of the basic aversions of the overwhelming majority of actual vertebrate earthlings. That is:

> What it is for $x$ to be a basic harm is for the overwhelming majority of subjects, as they actually are in this world now, to have a basic aversion to $x$.[49]

Basic benefits might be thought to count as such either (complementarily) in virtue of tending to elicit basic attractions in actual vertebrate earthlings, or rather simply in virtue of how they justify enduring certain basic harms. In any case, the term 'harm' seems to be typically used as a way of referring (rigidly) to a class of states of affairs including at least the basic harms. The normativity of the notion of harm is due to its connection to the notion of practical irrationality: what it is to be irrational is to fail to be averse to harms unless there is some compensating benefit.[50]

*2. Why can't IAP be vindicated just like Harm can?* A defender of IAP should now object that a parallel story can be told to justify IAP norms. For example, perhaps an account of the importance of scenarios-that-merit-Ingroup-loyalty can be given in terms of some response – say, reverence, or empathic disdain on behalf of the group – that the overwhelming majority of humans have to such scenarios. But such accounts are just not promising, for at least four reasons. First, the kind of agreement in response towards harms (and benefits)

Darwinian Dilemma for Realist Theories of Value', *Philosophical Studies* 127 (2006), pp. 109–66. Detractors include Jon Tresan, 'Question Authority: In Defense of Naturalism without Clout', *Philosophical Studies* (2010), pp. 221–38; and Stephen Finlay, 'Errors upon Errors: A Reply to Joyce', *Australasian Journal of Philosophy* 89 (2011), pp. 535–47.

[49] This is one version of Joshua Gert's account of basic harms. See Joshua Gert, 'Problems for Moral Twin Earth Arguments', *Synthese* 150 (2006), pp. 171–83, at 176. Gert characterizes an aversion as basic just in case the most appropriate response to the question, 'why are you averse to that item?' has no helpful answer other than 'what do you mean, why am I averse?'.

[50] Much more is said about this picture in Joshua Gert, *Brute Rationality* (Cambridge, 2004) and Joshua Gert, *Normative Bedrock* (Oxford, 2012).

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

is far more robust than that of any good candidate responses which could ground IAP norms. Several things are true of the basic aversion response to harms which are not true of candidate responses to IAP considerations. As HimpM suggested above, harm-aversion is shared by the overwhelming majority of vertebrate earthlings; it is triggered by the same list of things (pain, death, injury, etc.), and experienced to similar extents and valences; it is a well-defined, deeply motivating response, so ingrained in us that we can scarcely overcome it, and rarely have reason to. Parallel claims could not be truly made about the responses that correspond to IAP norms. Patriotic zeal, reverence for authority, disgust towards putative moral impurity: these are triggered by variegated lists of entities, in disparate extents and valences, varying tremendously in their natures and expressions across cultures, and in ways that wax and wane given our other commitments.[51] Second, harm has a tight connection to the fundamental norm of rationality, which IAP considerations do not. Someone whose action harms her without granting her a compensating benefit *ipso facto* acts irrationally, and *ipso facto* has reason not to perform this action. By contrast, someone who acts irreverently can be perfectly sane in confessing that she just does not see any reason to show the kind of respect or reverence expected of her in this case.[52] Third, as discussed in section II, welfare enjoys explanatory primacy with respect to IAP. Fourth, as discussed in section II, the intuitive and widely accepted exceptions to IAP norms suggest that welfare takes precedence, including cases in which IAP norms recommend lots of incredibly morally atrocious behaviour, from genocide to witch-burning.

*3. Must we accept IAP norms on their own terms?* Some, such as fans of rule-consequentialism, might object that even if IAP norms are heuristics for promoting welfare, we should still accept and be motivated by IAP norms for their own sakes. If we were not, the objection says, IAP norms would not succeed in promoting welfare. It might be that, for example, someone motivated to defer to Authority only when it is welfare-promoting is likely to be much less deferential, perhaps even when she should be, than if she simply defaulted to deference. This is parallel to how, on some consequentialist views, rules such as 'do not steal' and 'do not lie' need to take on independent psychological force. The thought is that exception-admitting rules such as 'do not steal, except when stealing increases overall welfare' will lead to lots of thefts due to cases where thieves think they won't be caught; but, the thought continues, many of these thefts actually

[51] On the disparate elicitors of disgust, see Kelly, *Yuck!*

[52] On this pair of points, see Gert, *Brute Rationality*, pp. 136–7.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

will be harmful, and the overall welfare will in fact decrease (perhaps even to the point that property conventions would be compromised). Similarly, it might be thought, in today's battlefields, offices and high schools, IAP norms should be followed for their own sakes, lest too much disloyalty, rank-breaking and promiscuity lead to unintended harms or even destroy those arenas' beneficial social structures.[53]

First, this objection seems partially rebuttable. In many contemporary contexts, it is plausibly false that – or at least unclear whether – intrinsic following of IAP norms will better promote welfare than following IAP norms only-when-they-clearly-promote-welfare. In contemporary diverse, socially evolving, large, democratic, liberal, nation-states, traditions of Ingroup, Authority or Purity often seem to diminish welfare when and to the extent that they encourage fighting against diversity, social accommodation, inclusion, proportional representation or free exercise of individual life-plans. It is most obvious where Ingroups or Authorities are dedicated to violently oppressing others (e.g. the Ku Klux Klan), but it is even clear where IAP structures (e.g.) unintentionally exclude minorities from various liberties, opportunities and privileges (e.g. exclude perfectly qualified women, freethinkers or transpeople from leadership positions). Within these contexts, it would be better if IAP norms were completely disregarded or at least thought to admit of exceptions. Breaking these IAP norms in these contexts will lead to fewer harms, and the social structures that are broken were not overall beneficial to begin with. It is not clear whether these contexts on the whole outnumber contexts where unqualified IAP norms would be useful.

Now to accommodate the objection. The best solution is to recommend that people be motivated by qualified versions of IAP norms. A good piece of public advice, which should intrinsically motivate people, is 'comply with Ingroup, Authority, or Purity traditions – except where this would be (unjustifiably) harmful'. Three relevant conditions are worth mentioning: (i) where an IAP norm recommends unjustified harm, people should not follow it and should disregard the IAP norm as a heuristic that fails in this case; (ii) where an IAP norm does not clearly promote welfare but also does not harm anyone (unjustifiably), people should follow it but consider it not to be *morally* important but only potentially so; and (iii) where an IAP norm does clearly promote welfare, they should consider it as a (contingently) moral rule, being intrinsically (but, we hope, cautiously) motivated by it. The qualification in (ii) will come from the public recognition that morality is centrally concerned with harm (roughly what we called HimpM

[53] Thanks to an anonymous reviewer for raising this helpful objection.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

above). This strategy can help coordinate activities in the sports arena, office or classroom without allowing the coordinated activity to begin diminishing welfare.

This IAP-norms-as-heuristics view admirably cleans up the normative landscape, makes sense of why we have IAP norms, and, by indicating when IAP norms admit of exceptions, helps us to avoid unnecessary moral dilemmas between welfare-promoting norms and distinct IAP norms.

### *III.5. Stubbornness*

Humans tend to maintain their original moral judgements even in the face of countervailing or clearly undermining defeaters. So testifies Haidt's research, and the surrounding literature, on the pervasive influence of intuition on moral judgement (intuition being conceived as quick, automatic, effortless, not always conscious evaluative mental states, including but not limited to emotional reactions).[54] Haidt highlights research which suggests that confirmation bias and motivated reasoning are ubiquitous features of human reasoning. This is so especially 'when self-interest, social identity, and strong emotions make them want or even need to reach a preordained conclusion'.[55] When people are asked to judge the permissibility, for example, of cleaning one's toilet with scraps of the national flag, or of consensual, protected incest between a pair of adults, subjects usually judge so as to condemn the act even though they must search hard for satisfying justifications, and are often unable to provide them.[56] These claims, if accurate, suggest that conservatives have such intuitions mostly because of their immediate intuitive or emotional reactions, and stick with them because of the human tendency to be 'intuitive lawyers' who justify their opinions through *post hoc* rationalizations.

Here it might be worried that liberals are similarly subject to debunking arguments in their characteristic moral judgements; after all, this theory holds that they, too, are 'intuitive lawyers'. However, the point of section II was that there is less need to provide any deeper justification for welfare-promotion norms. They help form an explanatory bedrock for morality, and even conservatives are committed to this claim. 'Stubbornness' about the moral relevance of

[54] See Haidt, 'Emotional Dog'. To be clear, I am not outright endorsing Haidt's 'Social Intuitionism', one problem with which is that it underemphasizes the role of intuitive rules in framing our moral judgements – as has been emphasized by such researchers as Susan Dwyer and John Mikhail. For discussion, see Ron Mallon and Shaun Nichols, 'Rules', *The Moral Psychology Handbook*, ed. John Doris (Oxford, 2010), pp. 297–320.

[55] Haidt, *Righteous Mind*, p. 81. See also the rest of Part I of that book.

[56] Haidt et al., 'Affect, Culture, and Morality'; see also Haidt, 'Emotional Dog'.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

Harm norms does not contribute to a potential debunking argument because such norms are not inaccurate heuristics for welfare promotion.

## IV. VINDICATING DURKHEIMIAN WELFARISM

Haidt is ambivalent about conservative morality. His work is mostly descriptive; he emphasizes how IAP norms are a useful and overlooked aspect of the norms about which humans moralize as a matter of (pre-) historical fact. Yet he does take a normative ethical position, endorsing a form of utilitarianism (at least for public policies in modern, diverse societies), and he admits this is a liberal view.[57] However, Haidt says the theory of the human good which we should adopt is 'Durkheimian' in holding that 'human flourishing requires social order and embeddedness' and that 'social order is extraordinarily precious and difficult to achieve', as well as holding that there are a plurality of values, some of which require a great deal of sympathy and empathy to understand.[58] However, Haidt offers no argument for this utilitarianism, and apparently thinks his value pluralism follows from his work's observation that IAP virtues and norms enrich and bind humans into persisting groups that they find important.

In broad agreement with Haidt, account and arguments given above begin to support a 'tolerant' form of welfarism, according to which the following claims hold:

(1) Welfare promotion (especially harm prevention) is the ultimate goal of all moral rules, ideals and virtues.
(2) IAP norms and virtues are in most human environments desirable yet overridable heuristics for promoting welfare for humans in their communities.
(3) 'Welfare' is to be construed broadly enough to cover certain desire-dependent interests, as well as various kinds of (un)freedom, ( (un)fair) opportunity, and (in)ability.
(4) There is a great deal of variety in which sorts of benefits (and other goods) it is rational for people to pursue, and a great deal of latitude regarding which sorts of benefits can rationally compensate for which kinds of harms.

(1) and (2) follow from what I have argued above. (3) follows from our characterization of welfare, and (4) is a plausible claim about value theory consistent with allowing people to consistently place high (moral) value on harms and benefits that may be constituted or created

[57] He calls utilitarianism a 'one-receptor system', i.e. one which takes harm to be the supremely important moral concept (*Righteous Mind*, p. 272).
[58] *Righteous Mind*, p. 272.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

by certain Ingroup, Authority or Purity-ritualistic relationships or practices. For example, where someone has an opportunity to sacrifice a great individual opportunity in order to be with the company of an ingroup, rationality seems to permit either choice, even if one of the two decisions may seem hard to justify from either an individualist or a collectivist perspective. The insight of this latitude is that welfarism can be flexible with respect to many questions of the value of relationships. For example, it can be *neutral* regarding questions like the following:

(a) whether healthy relationships with friends, family, members of ingroups, and superiors are irreducible aspects of people's welfare;
(b) whether community or other groups have interests that are irreducible to those of their members;
(c) how welfare should be distributed in the universe.[59]

Focus on question (a), noticing that a welfarist position can be relatively neutral regarding collectivist versus individualist values. There would be no room, and no point (here), to argue over whether friendship is intrinsically valuable, irreducibly (such that its value cannot, for example, be reduced to that of the mental states of friends). So, too, there would be no room to argue over whether *group* friendships (or affiliations or acquaintanceships) have similar intrinsic value. A Relaxed Form of Ingroup relationships (let us call it) merely views the Ingroup as a collection of group friends or affiliates or acquaintances; as with any friendship, the value of these is conferred by the relationship's candidate value-conferring constituent features – such as emotional bonds, mutual commitment to support in times of trouble, mutual positive interactions, a shared history, a sense of belonging, and (even) protection in a time of attack. Norms promoting these Relaxed conceptions of ingroups (and authority structures within them) do plausibly promote the general welfare. Think of groups pushing for rights-focused legal reforms: they unite and make symbols not for domination's but for compassion's sake.

This article's argument for MDT only casts doubt on Extreme Forms of Ingroup relationships: ones which recommend or require members to harm those inside or outside, i.e.: (i) to make significant, involuntary sacrifices for the group, though their personal interests be individually compromised; (ii) to display group symbols as a sign of competition

[59] Being open to (c) is one reason this tolerant welfarism is not consequentialist. I suggest it might be happily wedded with the view that the least advantaged should be made as well off as possible. The famous *locus classicus* of this view is John Rawls, *A Theory of Justice* (Oxford, 1971).

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press

against other groups; (iii) to envisage or act on a competitive attitude towards other groups that is unnecessarily destructive; or (iv) to dehumanize or harm people outside the ingroup. These groupthinkish mores are among the more destructive, tending to result in welfare-diminishing behaviours.[60] By contrast, whatever difficulties there may be in practice with promoting ingroup pride without also promoting hatred of outgroups, we should recognize that some ingroup concepts are more inclusive than exclusive or combative and lend themselves readily to the Relaxed conception of ingroups – e.g. honour societies and their symbols standing for scholarship and integrity; conceptions of world harmony and their world flags or vegan symbols.

Here one might worry that this is all too flexible: that concessions like (3)–(4) and (a)–(c) unhelpfully convert a debate about deontic requirements (etc.) into a debate about values. To be sure, these moves highlight open debates about a plurality of values, but they do not *merely* push back the debate. Harm and benefit are fairly well-understood concepts, as concepts go, in so far as we can produce substantive lists and characterizations of kinds of harms and benefits, at least of the most basic kinds of harms and benefits. Progress has been made because we can reasonably assess claims about IAP-type harms and benefits in light of our most plausible theories of harms and benefits; this is more manageable than when it was merely a question about which rules/norms apply. For example, if someone claims that anti-patriotism, secret mockery of elders or zoophilia are harmful, there can be productive debate about who is the patient of the harm (perhaps raising ontological debates about (e.g.) collective agency), what sort of harm it is (vulnerability-grounded? desire-grounded?) and (if vulnerability-grounded), whether the harm is an instance(s) of pain, death, disability, injury or illness of non-metaphorical sorts.

Far from oversimplifying ethics, the tolerant welfarist picture of morality allows us to make slow, steady progress on discerning which human norms are really relevant to the moral claims we are inclined to make, simplifying things just enough that substantive, yet difficult, debates about trade-offs in welfare can take place.[61]

jquigley@fsu.edu

[60] Adam Cureton has argued for a thesis that I take to be ambiguous between these. See Cureton , 'Solidarity and Social Moral Rules', *Ethical Theory and Moral Practice* 15 (2012), pp. 691–706.

[61] Many thanks to audiences at the 2012 Felician Ethics Conference and the 2012 Alabama Philosophical Society. Thanks to Michael Albert, David McNaughton, Tyler Paytas, Preston Werner and Chris Zarpentine for helpful comments. Thanks also to Scott Clifford, Josh Gert, Jon Haidt, Jesse Prinz and Tom Wysocki for helpful conversation or correspondence on this topic.

https://doi.org/10.1017/S0953820814000351 Published online by Cambridge University Press