

ARTICLE

Bain's Theory of Moral Judgment and the Development of Mill's Utilitarianism

Aaron Zimmerman 

University of California, Santa Barbara, USA
Email: aaronzimmerman@ucsb.edu

Abstract

In *Utilitarianism*, Mill defers to Alexander Bain's expertise on the subject of moral judgment to answer common criticisms of the creed. First, we do not blame people or label them immoral when they are less than ideal. Judgments of immorality are commonly reserved for substandard behavior, not suboptimal comportment. Second, we do not commonly insist on full neutrality in benevolence. Indeed, some philosophers argue that we are obliged to exhibit partiality, insofar as it is demanded by our roles as friends, parents, or children. My primary aim in this essay is to explicate Bain's theory of moral judgment and explain how Mill used Bain's psychological doctrines to inform his development of an indirect utilitarian moral philosophy, immune to the criticisms described above.

1. Utilitarian conceptions of obligation

In *Utilitarianism*, Mill defers to Alexander Bain's expertise on the subject of moral judgment to answer a common criticism of the creed. In developing his conception of utilitarianism, Mill defines goodness as pleasure or happiness and badness as suffering or unhappiness. Correlatively, according to Mill's often quoted statement of the principle of utility, "actions are right in proportion as they tend to promote happiness; wrong as they tend to produce the reverse of happiness" (1969a/1861, 210).

But Mill's conceptions of moral obligation and injustice are much more complicated than these preliminary analyses. If we ought to do good things and mitigate bad things, then (plugging in Mill's equations) it follows that we ought to promote happiness and diminish suffering. But it does not follow from this that we are *obligated* to produce happiness and avoid suffering; nor does this imply that it is *immoral* to fail to produce the most happiness and mitigate the most suffering that we can.

In fact, positing these obligations would pit utilitarianism against "common sense" morality as it was then and remains today. First, we do not blame people or label them immoral when they are less than ideal. Judgments of immorality are commonly reserved for substandard behavior, not suboptimal comportment. Common-sense moral thinking countenances a class of supererogatory actions we encourage but do not require.

Second, we do not commonly insist on full neutrality in benevolence (Jeske and Fumerton, 1997; Jeske, 2019). A parent is allowed to favor her young children over the children of strangers, and an adult child is allowed to favor her aging parents

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

over other elderly people, even if greater happiness or less suffering would result were they to adopt different objects of concern and affection. Indeed, some philosophers argue that we are obliged to exhibit partiality, insofar as it is demanded by our roles as friends, parents, or children. Susan Wolf is particularly forceful in her statement of the position.

It is absurd to suggest that morality requires one to care, or to act as if one cares, no more about one's own child than about a stranger's, or that it is immoral to go to the movies with a friend whenever more good could be done by working at a soup kitchen. (Wolf, 1992, 244)

Mill was well aware of both these items of common-sense morality. His recognition of common belief in the supererogatory will be quoted at length in our discussion below. Here Mill records Wolf's observations regarding the limits of impartiality.

A person would be more likely to be blamed than applauded for giving his family and friends no superiority in good offices over strangers, when he could do so without violating any other duty; and no one thinks it unjust to seek one person to another as a friend, connection or companion. (Mill, 1969a/1861, 243)

Should utilitarians reject these common-sense ideas as invalid because incompatible with their avowed first principle? Mill never explicitly addresses the possibility. Instead, the final chapter of *Utilitarianism* finds Mill trying to accommodate imperfection and partiality within his utilitarian conception of morality. The means he employs toward this end is a distinction between "immorality" understood as the failure to do what we are obliged to do, and "simple inexpediency," which often involves the mere failure to help others who have no acknowledged claim to our assistance.

According to Mill's observations on the matter, we (his audience) call an action "wrong" when we believe its agent ought to be punished, and we call a prospective course of conduct "right" when we believe the agent to whom it is presented ought to be compelled to execute it. These assertions, and the sentiments to which they give expression, are to be distinguished from encouragements and the desires implicated when we praise ideal forms of virtue.

We call any conduct wrong, or employ, instead, some other term of dislike or disparagement, according as we think that the person ought, or ought not, to be punished for it; and we say that it would be right to do so and so, or merely that it would be desirable or laudable, according as we would wish to see the person whom it concerns, compelled, or only persuaded and exhorted, to act in that manner. (Mill, 1969a/1861, 246; cf. Mill, 1989, 241–42)

Mill's use of "wrongness" and "rightness" to analyze moral obligation raises interpretive difficulties that continue to worry students and scholars of his philosophy. How can we reconcile the usage Mill here reports with the definitions of "rightness" and "wrongness" we have already extracted from Mill's prior statement of the principle of utility? If we were to blame (or deem blameworthy) all and only those actions which either promote suffering or fail to promote happiness, our usage of "right" and "wrong" as Mill here reports it, would conform to Mill's statement of the principle of utility. But, as Mill observes in the passage just quoted, we do not blame all failures to promote happiness. Nor do we blame every harmful act. It would therefore seem as though Mill must

abandon his principle of utility if he is to endorse the common-sense limitations on the application of “wrong” and “right” that he reports above. Put the other way around, it would seem as though Mill must reject common thought and speech if he is to retain his principle of utility, which would commit him to asserting an obligation we each have to maximize utility or at least promote it beyond what we commonly require of one another in daily life.

It is important to note at the outset that there is no textual evidence of any sort that Mill took the revolutionary course. He never asserts the onerous duties of benevolence advanced in our time by utilitarians like Peter Singer (1972) and Peter Unger (1996). Did Mill then abandon his first principle of morality? Or was Mill’s systematic treatment of morality ultimately inconsistent (Brink, 1997, 716; 2013, 102–3)?

I think a more charitable reading would interpret Mill as working with two distinct uses or notions of “wrong” in the passages in question. When one says that an action is “the wrong one to take,” one may mean no more than that there is a better action available, which will always be the case, from the perspective of a utilitarian legislator or moralist, when an alternative course would better promote utility (Mill, 1969/1861, 257–58). This usage gives us the sense of “wrong” Mill builds into the principle of utility, a sense which is relevant to, but not determinative of, immorality or impermissibility. For example, gambling, drinking to excess, and continuously procrastinating are all inadvisable insofar as these actions tend to detract from utility. (The same might be said of John’s token acts of gambling, drinking, or procrastinating.) But one can agree with Mill that these are the wrong things to do, and agree with Mill that they are “rendered” wrong by their negative effects on utility, without therein committing oneself to their immorality. One might think, for instance, that these acts involve inadvisable but permissible failures to achieve a utilitarian ideal or set of ideals.

In contrast, sometimes we call an action “wrong” to communicate that it is not only inadvisable, but immoral or impermissible, and so blameworthy or punishable in some way or other. This is a second sense of the term, which requires a distinct analysis of the sort Mill gives in the final chapter of *Utilitarianism*.

The question then for Mill is whether the requisite analysis can be made to square with the principle of utility. Folk rejections of perfectionism and full neutrality can be retained by the utilitarian if partiality and failures to perform intuitively supererogatory actions are not always blameworthy on utilitarian lights. What Mill must do, at a minimum, is provide an analysis of blameworthiness or liability to punishment which can be embraced by his audience without their rejecting his utilitarian equation of happiness with intrinsic value and his “proportionality doctrine” linking the objective rightness of an action with the degree to which it promotes happiness.

2. Sanctions utilitarianism

To my knowledge, Mill never explicitly provides a utilitarian analysis of blameworthiness to complete his analysis of immorality, which failure explains the continued debate on how Mill might have defined the concept had he done so. Perhaps the comment most often seized upon by those engaged in this speculative project is a remark Mill makes in a subsequent passage about the nature and enforcement of rights: “To have a right ... is ... to have something which society ought to defend me in possession of. If the objector goes on to ask why it ought, I can give him no other reason than general utility” (Mill, 1969a/1861, 250).

On this basis, one might interpret Mill as endorsing a “sanctions utilitarian” analysis of the concept of immorality or blameworthiness (Brink, 1997, 2013, 2018). Perhaps, for instance, Mill believed that an action is truly blameworthy just in case it ought to be blamed or punished. And perhaps he thought that we ought to blame or punish an action when doing so best promotes utility on the whole. We are to practice and encourage benevolence as a general matter. But actions are only judged wrong or immoral when we think they warrant punishment or reprimand. Since punishment and reprimand are actions in their own right, they are only warranted (from the utilitarian perspective) when they promote happiness or diminish suffering, whatever their epistemic credentials. To evaluate actions, then, we need only apply the principle of utility in two differing contexts: once when evaluating the action (type or token) being punished to determine whether it was or is harmful (or suboptimal) and so wrong in the first sense Mill identifies, and then a second time when deciding whether utility is best served by punishing it, which would render it immoral and so wrong in the second sense at issue.¹

There are, however, substantial interpretive obstacles to attributing this direct form of “sanctions utilitarianism” to Mill. First, Mill discusses cases in which we would justly punish people for their actions were it not for the disutility involved in meting out those punishments. If Mill equated an action’s being immoral with there being a utilitarian justification for punishing it, he should deem the relevant class of actions wrong (in the first sense of the term) but morally permissible (and so not wrong in the second sense). But Mill does not accept this classification of the actions he has in mind. Mill indeed insists that some harmful actions should not be punished by the state, but should only be admonished by the court of public opinion. And he allows that some of these actions should not be punished with a loss of reputation, but should be regretted and punished by the agent herself, who should feel guilty about the harm she has caused. But Mill goes further than this to suggest that some violations of duty should not be punished at all.

Duty is a thing which may be exacted from a person, as one exacts a debt. Unless we think that it might be exacted from him, we do not call it his duty. Reasons of prudence, or the interest of other people, may militate against actually exacting it; but the person himself, it is clearly understood, would not be entitled to complain. (Mill, 1969a/1861, 246)

And this suggests that an action is immoral, on Mill’s reckoning, just in case its agent is not “entitled to complain” about being punished for it, which would have Mill

¹The principle of utility is applied to action types when utilitarian legislators affix penalties to the violation of rules in advance, or utilitarian parents communicate the consequences of comparable violations to their children. In contrast, the principle would be applied to token actions were it used to calculate the penalty for a particular violation of a rule or set of rules. On my reading, Mill recommends the application of the principle of utility to types of action in prospect, but to action tokens when the rules conflict in their recommendations (1969a/1861, 206–7). The moral philosophy Mill advances is therefore neither act utilitarian nor wholly rule utilitarian in orientation. Mill’s analysis of the concept of impermissibility in particular shares the “indirectness” seized upon by ideal rule-utilitarian theories like Brandt’s (1979), but Mill eschews Brandt’s idealizations. We are obligated by those in a position to punish us to perform (or refrain from performing) certain *types* of action with *rules* (or less formal devices) which they use to frame their expectations. Morally impermissible conduct – the failure to shoulder an obligation – is, therefore, inevitably the violation of a rule. (See Mill, 1969a/1861, 220; Mill, 1972b/1867, 1234; and Mill, 1969b/1852, 173–80.)

acknowledging a whole class of genuinely immoral actions which cannot be punished by utilitarians for “forward-looking” reasons of instruction, deterrence, and the like.

It therefore seems, as an interpretive matter, that Mill did not equate the blame-worthiness of an action with the utility of punishing it. Instead, he countenanced immoral, blameworthy actions that we should neither punish nor blame. Mill simply maintains that in cases in which punishment for immorality would be infelicitous, the object of our punishment would not be entitled to complain were we to ignore the disutility of the proposed penalty to give him what he deserves.

A second obstacle to interpreting Mill as a sanctions utilitarian arises from his discussion of disagreements about the proper justification of punishment. Mill’s avowed aim in the final chapter of *Utilitarianism* is to show that justice is not “totally independent of utility” and is not a “moral standard per se.” As a means to this end, Mill argues, “not only have different nations and individuals different notions of justice, but in the mind of one and the same individual, justice is not some one rule, principle or maxim, but many, which do not always coincide in their dictates” (1969a/1861, 251–52). In support of this sub-thesis, Mill cites: (a) the views of those who argue that punishment cannot be justified to deter others from following the criminal’s example; (b) those who disagree by arguing that punishment can only be justified by these considerations; and (c) those, “like Mr. Robert Owen,” who argue that punishment is altogether unjust. Crucially, Mill does not take sides in this disagreement, but argues that “all these opinions are extremely plausible, and so long as the question is argued as one of justice simply, without going down to the principles which lie under justice and are the source of its authority, I am unable to see how any of these reasoners can be refuted” (252). If Mill were really a “sanctions utilitarian” in the sense we have defined, he would have taken sides in the debate. At the very least, he would have rejected Mr. Owen’s view and admitted the partial virtues of its alternatives. Instead, Mill treats each of these alternatives as an impoverished, partial conception. And this suggests that Mill envisioned an even less direct role for the principle of utility in fixing upon a correct standard of punishment than that advanced by our “sanctions utilitarian.” Indeed, this suggestion is borne out a few pages later, when Mill justifies schemes of *retribution* on utilitarian grounds.

The principle, therefore, of giving to each what they deserve, that is, good for good as well as evil for evil, is not only included within the idea of Justice as we have defined it, but is a proper object of that intensity of sentiment, which places the Just, in human estimation, above the simply Expedient. (Mill, 1969a/1861, 256)

Finally, when we turn to other texts, we find Mill embracing a division of labor between judges and legislators. Legislation is a moral art. An end is proposed and rules are crafted to achieve that end. This, Mill, says, is a context in which the principle of utility is rightly utilized to inform deliberation. Legislators should enact those rules they think will maximally benefit their constituents. In contrast, judges are supposed to apply and enforce the rules that emerge from this process. And Mill explicitly insists on the impropriety of a judge’s applying the principle of utility in this context.

The judge is not called upon to determine what course would be intrinsically the most advisable in the particular case in hand, but only within what rule of law it falls; what the legislature has ordained to be done in the kind of case, and must therefore be presumed to have intended in the individual case ... The legislator

is bound to take into consideration the reasons or grounds of the maxim; the judge has nothing to do with those of the law, except so far as a consideration of them may throw light upon the intention of the law-maker, where his words have left it doubtful. (Mill, 1974/1843, 944)

3. Mill's utilitarian defense of retribution

So what can we say about Mill's positive conception of immorality? Defined as it is in terms of blameworthiness, can it be reconciled with Mill's utilitarian theories of value and objective wrongness? Worries emerge when we try to analyze the conception of "entitlement to complain" Mill uses to advance his analysis of rights. Were the object of our punishments himself a sanctions utilitarian of the direct sort, he should complain of his punishment when and only when he thinks utility would be better served by his remaining unpunished. ("It won't do me any good," "I'm really harmless," and so on.) But the cases that Mill discusses above when advancing a penal criterion of immorality are precisely those in which a punishment is forgone because it would fail to serve utility of one sort or another. And Mill tells us that the violator he has in mind is *not* entitled to complain of the punishment contemplated for him in this circumstance. So the standards of entitlement Mill has in mind mustn't be set by an application of the principle of utility to the act of punishment itself.²

When, then, are we entitled to complain of our punishments? Common-sense answers are not limited to utilitarian evaluations of the particular punishment under consideration, nor evaluations of our punitive practices considered in their entirety, but include reasons extracted from standards of fairness which are "internal" to those practices themselves (Rawls, 1955). We are entitled to complain about a punishment if we didn't commit the act for which we are being punished, if we didn't know (or couldn't have known) that the act was harmful or subject to sanction, if we were forced into the act or unduly coerced to commit the act by another party, or if we couldn't have "done otherwise" for some equally valid reason. A person must have some minimal knowledge of what is expected of her, and the ability to meet these expectations, if she is to be justly blamed for her failure to do so. By limiting breaches of duty or obligation to those cases in which someone is not entitled to complain of punishment, Mill obliquely references a whole host of excusing conditions, conditions which are in turn grounded in conceptions of moral agency we have incorporated into our punitive practices. And it is natural to wonder whether it is coherent for Mill to advance a definition of immorality that depends on these ideas for its content. A pointed way to pose this worry would pit the view of justice Mill actually articulates in *Utilitarianism* against "sanctions utilitarianism" as we have defined it. Why shouldn't a utilitarian judge render the optimific decision? How can Mill justify his preference for a more indirect approach?

To justify the incorporation of backward-looking considerations within his theory of adjudication, commentators have interpreted Mill as embracing both "rule utilitarian" and "sanctions utilitarian" elements within a more complicated overall moral theory. For instance, according to David Lyons' influential analysis,

²Notice that these considerations cut across the type/token distinction. The individual violator is not entitled to complain about a just punishment, which judges ultimately forgo for utilitarian reasons. But so too a class of offenders is not entitled to complain of a schedule of just punishments, which legislators decide not to affix to an action-type for reasons of utility. Comparable observations hold for the informal analogues of legislation and punishment.

Mill had the following view. To call an act wrong is to imply that guilt feelings, and perhaps other sanctions, would be warranted against it. But sanctions assume coercive rules. To show an act is wrong, therefore, is to show that a coercive rule against it would be justified. The justification of a coercive rule establishes a moral obligation, breach of which is wrong. Someone punished under such a rule 'would not be entitled to complain' since he would truly be punished for a wrong done (not for doing something that is merely unpopular or illegal) ... But social sanctions may not always justifiably be imposed in the form of punishment for a past offence, since there may be overriding grounds for refraining from that further act. (Lyons, 1994, 55; cf. Brink, 2013, 101–3)

We thus get a utilitarian justification for a division of reasoning as between legislators and moralists on the one hand and judges and the court of public opinion on the other. Utilitarian legislators use the principle of utility to determine optimum laws to the best of their ability. A judge must reason from those laws to enact the utilitarian scheme chosen by the legislature and can only appeal to the principle of utility within the bounds set by the laws as she interprets them. The judge can justify this more limited use of the principle of utility within her judicial process by citing considerations of the overall utility established by the coercive system in which she is playing her part: principally, the advantages that accrue to the governed when they know the laws and have the opportunity to evade punishment by respecting them. As Mill says,

That a person is only responsible for what he has done voluntarily, or could voluntarily have avoided; that it is unjust to condemn any person unheard; that the punishment ought to be proportioned to the offence, and the like, are maxims intended to prevent the just principle of evil for evil from being perverted to the infliction of evil without that justification. (Mill, 1969a/1861, 257)

Of course, this implies that we would not be justified in “rendering justice” in a particular case were we to deem this act of retribution part of a coercive system we deem dysfunctional from our more foundational utilitarian perspective. As Piers Norris Turner (2015, 170) emphasizes, Mill consistently derides “rule worship” and blames participants in corrupt schemes for the larger wrong they conjointly create. Thus, to remain free from blame, judges must play their prescribed role in the scene without losing sight of the bigger picture.

It remains to be seen whether a similar analysis can be provided of our less formal moral system. Peter F. Strawson (1974), in particular, has drawn attention to a conflict between the utilitarian perspective Mill advocated and the reactive attitudes that continue to inform our responses to breaches of well-established obligations and clearly communicated expectations. But the idea would have each of us playing distinct roles as legislators and judges in a shared normative practice, which is indirectly grounded in considerations of utility. As legislators we communicate our expectations to others, prescribe their obligations, and propose terms of interaction. As judges, we enforce these obligations to the best of our ability with blame, distancing, reputational sanctions, and (often enough) violence. The social rules constituted by these expectations are objectively right, according to Mill, if their observance advances utility as intended. And the actions we sanction are genuinely immoral if they truly frustrate right expectations in a blameworthy manner. To determine the blameworthiness of a wrong action, we must not simply consider the utility of punishing its agent in a manner available to

us. At a minimum, we must also determine whether the actor is “entitled” to complain of this punishment, accepting the traditional array of explanations and excuses, so long as we can verify on reflection that honoring these excuses is an essential component of a beneficial scheme.

Thus, while blame and judgments of blameworthiness are not direct products of utilitarian calculation, and Mill does not think they should be, he insists that utility has played a grounding role in our thought and practice and that we should expand its role through reflection on those practices. First, though Mill overestimates the utility of morality, core components of our punitive practice do indeed promote utility and evolved because of this. Second, and perhaps more importantly, as a normative matter, Mill insists that we must have some assurance that a punitive practice promotes utility if we are to remain justified in playing our part in it. Intuitions of its intrinsic “fittingness,” which are given so much credence by the moral rationalists, are parochial and not to be trusted uncritically (Mill, 1969b/1852; Bain, 1875).

4. The Bain–Mill criterion of moral judgment

Though the passages I have subjected to analysis are canonical, and the interpretive difficulties I have raised are familiar to scholars of Mill’s work, a footnote, attached to these famous passages, is less commonly discussed. For after using a penal criterion to distinguish immorality from moral imperfection Mill asks us to: “See this point enforced and illustrated by Professor Bain in an admirable chapter (entitled ‘The Ethical Emotions or the Moral Sense’), of the second of the two treatises composing his elaborate and profound work on the Mind’ (Mill, 1969a/1861, 246 fn.).³ My aim in what remains of this article is therefore to explicate Bain’s theory of moral judgment and to see how it can be used to further develop our understanding of Mill’s indirect utilitarianism and the place of immorality within it.

In “The Ethical Emotions or the Moral Sense,” the chapter from the *Emotions* that Mill cites above, Bain does indeed advance a penal criterion for moral judgment and describe its relation to distinct if related evaluations. But Bain’s view is even more ambitious than Mill reports in *Utilitarianism*. Bain was not satisfied with the claim that belief in an act’s immorality is equivalent to the judgment that the act “ought” to be punished, which would mean analyzing one normative conceptualization in terms of another. Instead, Bain goes further to equate affective cognition of immorality with a feeling of disapprobation or resentment which includes or entails a willingness or “readiness” to punish the violator of a rule. The penal inclination is said to lend resentment, disapprobation, and allied emotions their “moral” character.

A moral rule in the strict sense is not an optional thing, but is enforced by the sanction of some penalty. It is true that practical ethics is made to embrace precepts of human virtue and nobleness, which undoubtedly deserve to be inculcated; but the compliance with those precepts constitutes merit and earns rewards, while the non-compliance does not entail punishment or censure. The question as to the morality of some line of conduct is – Does it inspire a feeling of disapprobation, as violating the maxims recognized to be binding? If so, the supposition is that the

³For instance, Lyons’ canonical (1994) interpretation of Mill on justice and obligation does not mention Bain. And Turner’s excellent analysis focuses on a correlative distinction Mill draws in his *Logic*. Mill’s avowed debt to Bain is overlooked.

same sense of duty that operates upon one's own self, and stings with remorse and fear in case of disobedience, should come into play when some other person is the guilty agent. The feeling that rises up towards that person is a strong feeling of displeasure or dislike, proportioned to the strength of our regard to the violated duty. There arises a moral resentment, or a disposition to inflict punishment upon the offender. It is the readiness to punish, that forms the criterion of moral disapprobation, or marks the boundary between a moral sentiment and an allowable difference of opinion. (Bain, 1875, 291)

So a preliminary question concerns this divergence between Bain and Mill. Is blame itself the criterion of moral disapprobation or are judgments of blameworthiness better suited to play this role? Though the question might seem trivial, it is crucial to a host of issues that continue to divide moral psychologists today. For instance, if blame and other forms of punishment are sufficient for moral judgment, as Bain insists, non-human animals may enact moralities, as many social animals punish one another. But if, as Mill intimates, judgments of blameworthiness are necessary, morality is probably an exclusively human affair, as we have no evidence that any of the other animals protest their punishments as unwarranted. (Nor do we have evidence of a non-human animal defending a punishment as a fitting response to the action at issue.) Indeed, the kinds of protest and defense in which judgments of warrant figure are discursive in nature; they are inaccessible to animals who lack sentential language. Second, if, as Bain insists, a subject must be willing to punish herself for an action if she is to be credited with judging it immoral, moral judgment will turn out to be “intrinsically motivating” in a straightforward sense, warranting an “internalist” conception of moral thought. But if, as Mill maintained, a subject must only judge herself worthy of punishment to judge her action immoral – and this judgment is in turn analyzed as a discursive or conceptual state, which is not itself painful, and which does not itself include or in some other way necessitate painful emotions like regret and remorse – the motivation to act morally might be thought to be “external to” that subject’s judgment.⁴ Thirdly, if Bain is right, and punishment is criterial for cognizing something as immoral, we can uncover a community’s moral rules by observing their punitive practices. But if Mill is right, we must instead focus on that community’s discursive or justificatory practices. Finally, if Bain is right, we can uncover an individual’s morality by describing the maturation of her punitive practices. But if Mill is right, we must instead track her capacity to evaluate punitive practices by formulating views about which actions ought to be punished and why. Since Bain’s and Mill’s definitions of moral judgment hold different implications for theories of the ontogeny and phylogeny of moral rules, and different implications for theories of moral knowledge and motivation, we cannot overlook this intramural dispute.

To begin our evaluation of the discrepancy between the two analyses, we must unpack the meaning Bain attaches to “criterion” when he writes of “the criterion of moral disapprobation.” The term has Platonic connotations, but nowhere in his extensive analyses of immorality does Bain assert necessary and sufficient conditions for moral condemnation. Instead, Bain employs sketches of the sort quoted above to outline paradigmatic cases in which we experience and judge things immoral and then singles out the penal component present in all these cases as “criterial.”

⁴Cf. Mill (1972a/1859, 649), where Mill argues that “I ought to X” has two distinct meanings depending on whether the speaker thinks she would feel guilt were she to fail to X.

Punishment “lends” a rule its moral character and “marks the boundary” between immorality and allowable differences, but Bain is clear throughout that there is always more to moral judgment than punishment. Though criterial, punitive practice is insufficient.

In fact, though Darwin accused Bain and Mill of excessive empiricism, Bain is explicit in attributing components of our moral thinking to *innate* psychological traits or capacities which have nothing to do with punishment.⁵ But because Bain thinks of the “readiness to punish” as necessary for moral judgment, he concludes that sympathy, love, and the rest of our innate moral endowment are also involved in forms of thought that are not distinctively moral in character. It is only when the emotions and judgments issued from these innate capacities are conjoined with an interiorization of authority that distinctively moral cognition is possible. Bain rejects nativist accounts of morality on this basis, as the criterial conception of authority cannot be inherited biologically.

In declaring against the intuitive character of the [moral] sentiment, I freely avow that certain powers belonging to us at birth are indispensable to the growth of our moral feelings; which powers, one or all, may have been developed through the cumulative experience of past generations. Into the moral sentiment there enter, first and most conspicuously, the ordinary action of the Will, prompting us to avoid pain and court pleasure; secondly, the power of Sympathy, or disinterested impulses; and lastly, the operation of our leading emotions, as Fear, Love, Anger, and others. All these are primitive or inborn powers of the mind, and as such are open to be accounted for by the hypothesis of development [i.e. biological evolution]. The finishing stroke, in my view, is due to education under Authority; which constitutes the moral sentiment a distinct and peculiar phenomenon, different from all the other exercises of Will, Sympathy, and Emotion, or any compounds of these. (Bain, 1875, 57)⁶

Perhaps, then, we should read Bain and Mill as equating distinctively moral judgments with blame or judgments of blameworthiness directed at people or institutions for the violation of social or interpersonal norms. As Bain says, the penal element “turns a rule into a law.”

In the paradigmatic case in which a person thinks that she has acted immorally, she feels the sting of remorse and experiences a kind of fear when she contemplates performing the act again. And though Bain denies its innateness, he claims that the mechanism responsible for this reaction has its origins in early development: “There are other impulses in the mind that induce a regard to the feelings of others; but with respect to duty, strictly so-called, the infant conscience is the linking of terror with forbidden actions” (Bain, 1875, 286). At the same time, “A sentiment of love or respect

⁵“Mr. Bain and others believe that the moral sense is acquired by each individual during his lifetime. On the general theory of evolution this is at least extremely improbable. The ignoring of all transmitted mental qualities will, as it seems to me, hereafter be judged as a most serious blemish in the works of Mr. Mill” (Darwin, 1889, 98).

⁶It would also be uncharitable to interpret Bain’s penal criterion as supplying sufficient conditions for moral judgment, as this would saddle Bain with a false or radically unintuitive theory. If common thought is to be trusted, there is such a thing as non-moral blame, as when an athlete is blamed for her team’s loss without imputation of immorality.

towards the person of the superior, infuses a different species of dread from what we have just supposed, the dread of giving pain to a beloved object. Sometimes this is a more powerful deterrent than the other. We call it a higher order of conscience to act from love than to act from fear” (Bain, 1875, 286). But Bain does not limit moral cognition to fear of punishment or loss of love and the need for parental approval. As Aristotle emphasized, the properly habituated child will soon learn the point of acting morally, which, if the morality imposed by the child’s handlers is at all justifiable in liberal terms, will consist in the role the imposed rules play in promoting happiness and diminishing suffering.

When the young mind is able to take notice of the use and meaning of the prohibitions imposed upon it, and to approve of the end intended by them, a new motive is added, and the conscience is then a triple compound, and begirds the action in question with a three-fold fear; the last ingredient being paramount, in the maturity of the sympathies and the reason. (Bain, 1875, 286; cf. p. 288)

In contrast, when we cognize someone’s else’s activity as immoral, we dislike them for something they’ve done, which emotion Bain labels “disapprobation.” Disapprobation and allied forms of resentment are supposed to involve (as a component element) some disposition to punish the offending agent in one or more of the various ways Mill would go on to accept as criterial: reputational sanctions, distancing, shunning.

The powers that impose the obligatory sanction are Law and Society, or the community acting through the Government by public judicial acts, and apart from the Government, by unofficial expressions of disapprobation and the exclusion from social offices. The murderer and the thief are punished by the law; the coward, the adulterer, the heretic, the eccentric person, are punished by the community acting as private individuals, and agreeing by consent to censure and excommunicate the offender. (Bain, 1875, 264)

To say that punishment is “criterial” for moral disapprobation is to say that it is the component of that phenomenon that we can use to distinguish it from other aspects of our psycho-social interactions.

With this analysis of Bain’s penal “criterion” in hand we can now return to examine its divergence from the account of immorality Mill articulates in the final chapter of *Utilitarianism*. The extent of the divergence depends on how Mill conceptualized the judgment that someone “ought” to be punished for a given course of action. The move to normative judgments about punishment (in contrast with dispositions to punish) might be justified on the basis of examples confronted by sentimentalists in the literature with which Mill and Bain were both familiar. For example, Hume posited forms of “emotionless blame” grounded in counterfactual reasoning.

We blame equally a bad action, which we read of in history, with one perform’d in our neighbourhood the other day: The meaning of which is, that we know from reflection, that the former action wou’d excite as strong sentiments of disapprobation as the latter, were it plac’d in the same position. (Hume, 2000, T 3.3.1.18; SBN 584)

But do we *really* blame ancient criminals for their ancient crimes (as Hume asserts) or do we merely judge them blameworthy? If some sort of negative phenomenology is essential to blame, the latter analysis is more apt. And since, as an intuitive matter, we do believe in the immorality of past crimes, this classificatory decision would seem to favor Mill's analysis over Bain's. The judgment that a long dead criminal deserved punishment underwrites one's belief in that criminal's immorality. Indeed, it is hard to see how we can join Bain in asserting that a "readiness" to punish a miscreant is necessary for belief in his immorality, when the object of our belief is no longer around to be punished. To accommodate this sort of case, Bain would have to posit unactualizable dispositions to punish that fail to manifest as sentiments. When, for example, we judge the antebellum slaveholder immoral, we acquire a disposition to punish him, where this means we would punish him were we to have the chance, a counterfactual that can be true even if we know the chance is impossible.

A similar *prima facie* obstacle for Bain's analysis arises when we allow that a friend or loved one has acted immorally but we find it impossible to feel blame towards them and have no desire to see them punished. On Mill's account, we can judge a loved one immoral so long as we believe they deserve blame we cannot feel and deserve punishments we cannot bring ourselves to deliver. Again, to secure these same results, Bain must posit dispositions to blame and punish that are prevented from actualization by the judge's allegiance with the criminal. Feelings of love overwhelm the causal efficacy of a real (if unactualized) readiness to punish.

But there are also cases which are more amenable to Bain's analysis than Mill's. As utilitarian critics of Victorian morality, both Mill and Bain allow that we often blame people for harmless actions for which they should not be blamed. In many of these cases, we incorrectly judge the agent blameworthy because we are misled by intuitions of the "intrinsic fit" between our blame and the deed we have blamed. But in some cases, we blame without considering whether the deed indeed deserves blame. And in a range of these cases, were we to consider whether the action really deserves blame, we would judge that it does not. When we blame without judging an action blameworthy, does our blame enable a judgment of immorality? A positive answer favors Bain's analysis over Mill's.

Consider, for example, Jonathan Haidt's (2001) well-known experiments. Subjects are handed stories of adult siblings who engage in incest without any negative consequences. Contraception is used, the relationship is strengthened, no one finds out, and so on. Similar stories are recited about subjects acting in "unclean" ways without consequence or privately desecrating flags and other symbols of religious and political identity. Predictably, many subjects say they find these actions immoral even when they cannot supply any substantive reasons or arguments to support their judgments. Instead of allowing exceptions to more general prohibitions on harmful or disrespectful actions, these subjects double-down on their moral views and claim to "just know" that the actions in question are "not OK": a phenomenon Haidt has labeled "moral dumbfounding."

One fairly natural interpretation of moral dumbfounding is that people reactively moralize incest, bestiality, and so on without first making any judgment one way or another as to whether their reactions are justified or warranted. The "test cases" are those in which someone blames the siblings' act of incest, and is ready to punish them with reputational sanctions, but then *abandons* this frame of mind when she considers whether incest deserves punishment in the case on hand. At the outset of this process, the subject blamed the siblings without deeming their actions blameworthy

in any substantive sense. Nevertheless, she sincerely accused the incestuous siblings of immorality. Shouldn't we conclude then, with Bain, that she believed what she said? But if this interpretation is right, Mill's criterion imports too much sophistication into the nature of moral judgment. People can judge others immoral without thinking one way or another about whether the actions they abhor have qualities that would render such a judgment warranted. Sometimes, all it takes is disgust with what's been done or the party who's done it.

To handle both sorts of case, we need a disjunctive criterion. So let us coin a new view, "the Bain–Mill criterion of moral judgment," according to which: (a) the disposition to blame or punish some person or institution for an action they are supposed to have executed, or (b) the disposition to judge some person or institution worthy of blame or punishment for some such action, is necessary for belief in the immorality of what's been alleged. To be clear, as a historical matter, Bain proposed a penal criterion for moral judgment, and Mill altered that criterion in substantive ways that he failed to acknowledge. I have crafted the Bain–Mill criterion to incorporate the advantages of Mill's unacknowledged alterations, without shouldering the controversial intellectualist commitments these alterations would otherwise entail.

5. Imperative duties and sentimental proscriptions

It is important to note that the Bain–Mill criterion is psychological, not explicitly normative. As Bain conceptualizes the matter, the nature of moral judgments, rules, and sentiments is a matter to be settled by science, insofar as each society has a morality that can be uncovered through careful observation and analysis. But we go beyond the sciences of moral psychology and ethology (or sociology) when we consider "the reasons or considerations requiring each one to abstain from the performance of certain actions, and to concur in a common prohibition of them," as this is a matter for the "Theory of Morals" (Bain, 1875, 265). When an intellectual engages in the Theory of Morals, she is attempting to either critique or justify a morality (in the sense of "morality" induced by adoption of the Bain–Mill criterion), where we do not or need not assume that the morality we are currently enacting is the morality we will have after we have developed a sociological description of our practices and subjected it to critique. Importantly, critique and justification are not elements of moral psychology or any other science. They are forms of art.

In remarking upon these different views of the origin of moral distinctions, we must bear in mind that it is one thing to inquire what has been the motive for setting up the rules that we find existing in any community, and another thing to settle the motive that we think ought to govern the imposition of those rules. To explain historically the rise of institutions is different from the endeavor to settle the best principles for modifying the old, or for forming new. It may be that some portions of the existing morality have been generated by considerations or motives that we dissent from, although we cannot deny that such motives have operated to produce the result. (Bain, 1875, 265)

The distinction between scientific conceptions of the origins of a given person's morality and the normative arguments we might give for the validity or invalidity of that morality is easy to keep in place when moral judgment is defined as Bain and Mill define it. A person or group of people may be disposed to punish themselves or

others for good reasons, bad reasons, or no reasons at all. Either way, the schedule of punishments they enact is a morality in the sense Bain defines, and we can find their moral judgments among those states of mind motivating each to play her part in the scheme: the disapprobation, the guilt, the derogation, and ostracization. These are moral sentiments even if the judgments that express them are ill-conceived, and the punishments to which they give rise are harmful and undeserved. Of course, this way of thinking lies at odds with those who define moral judgments as the products of sympathetic engagement with the feelings of others, or contractualist reasoning, or egalitarian impulses, or anything else we might associate with the truth or validity of a moral judgment. But this is intentional. Bain's analysis is crafted to reject any link between the definition of moral judgment, and the truth, validity, or reasonableness of judgment so defined. Just as legislatures can enact wholly bad laws, communities can realize wholly bad moral norms.

A particular line of conduct is so intensely disliked for some reason, *or for none*, that we are prepared to resent it and to hinder the performance with all our might. A majority of our fellow citizens takes the same strong view, and actually employ their power as the majority, to prevent it absolutely; whereby a new article in the moral code of that community is set up. As a matter of course, any one committing the forbidden deed is disapproved of, and handed over to be punished. (Bain, 1875, 291–92, emphasis added)

Nevertheless, Bain prefaces his scientific account of the cultural evolution of moralities with a series of arguments he levels against the leading normative moral theories of the time. Divine command theory is rejected out of hand because it hasn't even secured consensus among Christian philosophers. The rationalist moralists (Cudworth, Wollaston, Clarke, and Price) are dismissed with a Humean critique: we calculate and reason to apply laws and find out the means to our ends, but reasoning is an insufficient guide to the ends themselves. Kant is allowed to be right about the need for generalized laws within a community, but his categorical imperative is inadequate as "it does not say what ends should be sought by this uniformity of procedure." The same problem besets Adam Smith's "impartial spectator": the perspective of a spectator is inferior in some ways to that of a participant, and even a perfect judge renders imperfect justice when she applies bad laws. Hobbes is right when he says, as a scientist, that "Morality is an Institution of Society, maintained by the authority and punishments of Society." But Hobbes' normative claim that society "made, or should have made, their last will and testament in favour of some single despotic ruler" was "decided against him by the events." Self-interest is not an acceptable moral guide because we have irreducibly other-regarding desires, interests, and concerns. And the moral sense theory, as instanced by Whewell in his *Elements of Morality*, assumes universal standards that are not to be found (Bain, 1875, 265–70).

It is only at this point, when arguing against Whewell's hypothesized moral sense, that Bain finally returns from the Theory of Morals to the science of morals. If the analogy with our five senses is taken "strictly," moral sense theory predicts relatively universal (or species-typical) moral judgments. And though there are, Bain allows, certain rules that are found in all human moralities, the explanation for them is not a common sense, but the role enforcement of these rules plays in the maintenance of communal ties.

Doubtless, if the sad history of our race had been preserved in all its details, we should have many examples of tribes that perished from being unequal to the conception of a social system, or to the restraints imposed by it. We know enough of the records of anarchy to see how difficult it often is for human nature to comply in full with the social conditions of security; but if this were not complied with at all, the result would be mutual and swift destruction ... If the sense of common safety were not sufficiently strong to constitute the social tie of obedience to some common regulations, society could not exist to tell the tale of an exception to the universality of a common standard of right. (Bain, 1875, 271)⁷

Of course, a society's morality is not exhausted by the minimal schedule of punishments necessary for its persistence. As we've seen, Bain recognizes a second class of moral rules "of purely Sentimental origin, having no reference, direct or indirect, to the maintenance of the social tie" (1875, 272). The rules regulating sex, diet, and dress are cited as the most perspicuous examples. There is uniformity in the first class of rules, because societies face similar obstacles to their persistence and elect common remedies. But "in the rules founded on men's sentiments, likings, aversions, and antipathies, there is nothing common but the fact that some one or more of these are carried to the length of public requirement, and mixed up in one code with the imperative duties that hold society together" (Bain, 1875, 273).

It is only after Bain has this genetic, sociological distinction in hand – between the "imperative duties" on the one hand, and those moral rules generated by mere sentiment on the other – that he examines utilitarian moral theory and describes how it can be defended as a normative Theory of Morals.

Some limit must be set to the principle [of utility], for it is obvious that we do not make everything a moral rule that we consider useful ... Jeremy Bentham wrote many useful books, but not because of his being obliged to do so under a high moral sanction ... A distinction must therefore be drawn between Utility made compulsory, and what is left free. (Bain, 1875, 276)

And it is at this stage that Bain directs the reader to his *The Moral Philosophy of Paley* (1852), where he provides two utilitarian arguments for the restriction of duty to that which is necessary for safety and security. The first is an acknowledgement of the limitations of the legislator and moralist. Punishment is the tool they employ to reach their ends, and it has a limited and uncertain utility: "The resources of the moralist are not omnipotent; they are extremely limited, and corresponding limits ought, therefore, to be set to his aims" (Bain, 1852, 90–91). As an epistemic matter, moralists can be more or less certain that security is necessary for happiness and they can therefore safely impose the "imperative duties" by punishing actions which undermine it. But we must admit that "by attempting to increase the share [of happiness] that comes from without, [we] may possibly diminish the share that accrues from within" (Bain, 1852, 90) and this justifies great caution when using punishments to secure further ends. The

⁷Bain's rejection of innate moral beliefs and knowledge remains controversial. Bain compares morality to architecture and language (1875, 272), which he thinks of as natural to humans, but not innate in the sense he has in mind. For support of this empiricist outlook see Prinz (2008). Against it, see Rawls (1971), Mikhail (2011), and Dwyer (1999), who all model our sense of justice on Noam Chomsky's "universal grammar." Cf. Zimmerman (2013).

moralist's distinction between obligation and supererogation is warranted by real limitations in our knowledge and power. It is too risky to moralize behavior when we do not know whether our admonitions can secure these further ends and we do not know whether our punishments will have negative effects beyond the pain involved in their execution. So we should "command" respect for basic rights on pain of punishment, but we should only "encourage" benevolence beyond these limits.

The second route Bain articulates is more speculative and Panglossian insofar as it assumes as a matter of fact what can only be determined by an evolving social science: that more happiness is promoted when we enact a limited schedule of social and legal punishments than would be enjoyed were we to institute penalties for each and every forgone opportunity to promote utility. This assumption lies implicit in Mill's *Utilitarianism*, but it is clearly articulated and tacked on to Bain's analysis as a "tenable" conjecture.

There is ... one mode of conceiving the greatest happiness principle, which comes to the same results as those above given. We might, for example, lay down limits to obligation, and claim the utmost possible scope to human liberty, on the ground, that in this way we promote most effectually the happiness of mankind. This is a perfectly fair and tenable position. We thus maintain the greatest happiness of the whole as the ultimate standard, and yet set bounds to men's interference with each other. (Bain, 1852, 91)

6. Mill's errors

We have seen how Bain introduced a penal criterion for judgments of immorality and used it to define an indirect form of this venerable moral theory, introducing innovations which were adopted by Mill in his now famous work on the subject. But we have also seen that the two thinkers were not exactly of one mind. In fact, in the later editions of *The Emotions*, Bain appended two explicit criticisms of Mill's *Utilitarianism* to his analysis.⁸ First, in the penultimate chapter of *Utilitarianism*, Mill argues that happiness is the only thing we desire for its own sake, and Mill infers, on this basis, that happiness is the only thing desirable for its own sake. In the subsequent decades, many critics would take aim at this reasoning, but Bain was one of the first to challenge its premise. The saint needn't include virtue in his conception of happiness, nor need the miser find happiness in the mere possession of money. Instead, what were originally conceived as means to happiness may become ends in themselves distinct and opposed to happiness.

The theory of disinterested action, in the only form I can conceive it, supposes that the action of the will and the attainment of happiness do not square throughout. There is an exceptional corner, not very large in the mass of men, where motive and happiness come into conflict. (Bain, 1875, 296)

Second, and relatedly, Bain criticizes Mill for underestimating the role that mere sentiment has played in the origins of our moral codes. Mill gives too great a weight to utility as a cause. In particular, Bain points to cases in which utilitarian rationalizations are provided for sentimental aversions.

⁸For reference, the first edition of Bain's *The Emotions and the Will* was published in 1859, Mill's *Utilitarianism* was first published two years later (1861), and the second and third editions of Bain's *Emotions* appeared in 1865 and 1875 respectively.

It is sometimes said in defence of the antipathy of the white population of the United States as to the persons of the free blacks, that this springs out of a wholesome dread of deteriorating the whole breed by mixture with an inferior race; to which the reply is, that the consideration of bringing forth a vigorous progeny has never been a ruling consideration in any known community, excepting ancient Sparta. It is usual enough to advance utilitarian *pretexts* for sentimental requirements, but the emptiness of such pleas is usually apparent. (Bain, 1875, 279–80 fn., emphasis in original)

Mill's *Utilitarianism* is a polemic. Its ultimate aim is to promote and defend a moral philosophy, not to describe the origins and function of our existing morality. Perhaps because of this orientation, Mill misuses science to support his normative aims, and falls afoul of Hume's Law, which commands theorists to distinguish fact from value in their analyses.

Nowhere is this more evident than in Mill's treatment of justice in the final chapter of *Utilitarianism*, where Mill attempts to provide a debunking explanation of deontological or rights-based views of morality. It is in this context that Mill famously writes that the sentiment of justice does not involve considerations of utility or expediency, but "whatever is moral in it does" (Mill, 1969a/1863, 248): "The sentiment, itself, has nothing moral in it, what is moral is, the exclusive subordination of it to the social sympathies so as to wait and obey their call" (Mill, 1969a/1863, 249). Were he utilizing the Bain–Mill criterion, Mill would allow, with Bain, that an interpersonal sentiment (or "reactive attitude") is moralized by an individual as soon as she punishes those who evoke it; and that once they are widespread, the induced social sanctions constitute a moral rule. The rule may be a bad one when evaluated from a utilitarian perspective, and Mill does much in the preceding pages to recommend that perspective. But however attractive, utilitarian evaluation is, by Mill's own account, a matter of art, not science (Mill 1969a/1863, 206). To respect Hume's Law, the critique of retributive institutions must be distinguished from a scientific explanation of their origins. The adoption of the utilitarian perspective is often irrelevant to the genesis of the particular punitive structures in play. Indeed, we should expect people who are not utilitarians to justify their blame, resentment, and disapprobation on non-utilitarian bases, to provide discursive support for feelings they have already "moralized" in the sense of that term induced by the Bain–Mill criterion.

7. Conclusions

Bain's criticisms are valid. As a consequence, Mill's arguments in the final chapter of *Utilitarianism* fail to undermine Kantian, rights-based, or deontological moral views, which assign respect and allied values a foundational role. It is reasonable to believe that victims of a crime have a standing from which they can demand justice without being accused of seeking revenge. And it would be unreasonable for Mill to insist that victims justify their demands with an appeal to the utility of the laws or mores that were violated and the role that punishments of the desired sort play in maintaining these laws (cf. Darwall, 2013). On what other grounds might victims justify their demands for punishment? Must we join Whewell in donning that "pitiful intellectualist trinket": an appeal to the "intrinsic fittingness" of punishment to crime? No. Just as utilitarians treat happiness as an ultimate value by desiring it for itself upon reflection, so too can egalitarians appeal to their dignity as an ultimate value by demanding respect without further justification.

But utilitarians can concede the importance of mutual respect in contemporary thinking about rights and justice – and they can allow Kantians the freedom to justify punishments with considerations of mutual respect rather than utility – without countenancing foundational judgments of justice or mutual respect. Indeed, Bain already achieved this result in 1875, by carefully distinguishing moral psychology from moral theory, and by similarly distinguishing our punitive practices from the use of utilitarianism’s first principle to critique those practices. We can use considerations of utility to improve rather than replace justice, just as Mill thought.

In sum, the Bain–Mill analyses provide utilitarians with attractive ways to think about immorality, the violation of rights, and the place these concepts occupy in enlightened moral thinking. Utilitarianism is not, as Mill claimed, the only rational morality to adopt, and the final chapter of *Utilitarianism* fails to “debunk” its rationalist target. But the indirectly utilitarian conception of immorality that Mill sketched to conclude his masterwork is deep and cogent. And though Bain’s contributions to utilitarianism have been largely forgotten, much of that philosophy’s cogency can be attributed to Bain’s innovative moral psychology.⁹

References

- Bain, Alexander**, 1852. *The Moral Philosophy of Paley: With Additional Dissertations and Notes* (Edinburgh: William and Robert Chambers).
- Brandt, Richard B.**, 1979. *A Theory of the Good and the Right* (Oxford: Clarendon Press).
- Bain, Alexander**, 1875. *The Emotions and the Will*, 3rd edition (London: Longmans Green and Company).
- Brink, David**, 1997. Review of David Lyons, *Rights, Welfare, and Mill’s Moral Theory*, *Philosophy and Phenomenological Research*, 37.3: 713–17.
- Brink, David**, 2013. *Mill’s Progressive Principles* (Oxford: Oxford University Press).
- Brink, David**, 2018. Mill’s Moral and Political Philosophy, *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), ed. by Edward N. Zalta, <<https://plato.stanford.edu/archives/win2018/entries/mill-moral-political/>>.
- Darwall, Stephen**, 2013. *Honor, History and Relationship: Essays in Second-Personal Ethics, Volume II* (Oxford: Oxford University Press).
- Darwin, Charles**, 1889. *The Descent of Man and Selection in Relation to Sex*, 2nd edition (New York: D. Appleton and Company).
- Dwyer, Susan**, 1999. Moral Competence, in *Philosophy and Linguistics*, ed. by Kumiko Murasugi and Robert Stainton (Boulder, CO: Westview Press), pp. 169–90.
- Haidt, Jonathan**, 2001. The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment, *Psychological Review*, 108: 814–34.
- Hume, David**, 2000. *A Treatise of Human Nature*, ed. by David F. Norton and Mary J. Norton (Oxford: Oxford University Press); originally published 1739.
- Jeske, Diane**, 2019. Special Obligations, *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), ed. by Edward N. Zalta (ed.), <<https://plato.stanford.edu/archives/fall2019/entries/special-obligations/>>.
- Jeske, Diane and Fumerton, Richard**, 1997. Relatives and Relativism, *Philosophical Studies*, 87.2: 143–57.
- Lyons, David**, 1994. *Rights, Welfare and Mill’s Moral Theory* (Oxford: Oxford University Press).
- Mikhail, John**, 2011. *Elements of Moral Cognition: Rawls’ Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment* (Cambridge: Cambridge University Press).
- Mill, John Stuart**, 1969a. Utilitarianism, in *Collected Works of John Stuart Mill, Volume 10: Essays on Ethics, Religion and Society*, ed. by J. M. Robson (Toronto: University of Toronto Press), pp. 203–60; originally published 1861.

⁹For invaluable comments on earlier drafts of this article I am grateful to Stephen Darwall, Dale Miller, and an anonymous referee for this journal.

- Mill, John Stuart**, 1969b. Whewell on Moral Philosophy, in *Collected Works of John Stuart Mill, Volume 10: Essays on Ethics, Religion and Society*, ed. by J. M. Robson (Toronto: University of Toronto Press), pp. 165–202; originally published 1852.
- Mill, John Stuart**, 1972a. Letter to William George Ward, November 28, 1859, in *Collected Works of John Stuart Mill, Volume 15: The Later Letters of John Stuart Mill 1849–1873*, ed. by F. E. Mineka and D. N. Lindley (Toronto: University of Toronto Press), pp. 646–50.
- Mill, John Stuart**, 1972b. Letter to Henry S. Brandreth, Feb 9, 1867, in *Collected Works of John Stuart Mill, Volume 16: The Later Letters of John Stuart Mill 1849–1873*, ed. by F. E. Mineka and D. N. Lindley (Toronto: University of Toronto Press), p. 1234.
- Mill, John Stuart**, 1974. A System of Logic, Ratiocinative and Inductive, in *Collected Works of John Stuart Mill, Volume 8: A System of Logic, Ratiocinative and Inductive, Books IV–VI, Appendices*, ed. by J. M. Robson (Toronto: University of Toronto Press); originally published 1843.
- Mill, John Stuart**, 1989. James Mill's Analysis of the Phenomena of the Human Mind, in *Collected Works of John Stuart Mill, Volume 31: Miscellaneous Writings*, ed. by J. M. Robson (Toronto: University of Toronto Press), pp. 93–253.
- Prinz, Jesse J.**, 2008. Resisting the Linguistic Analogy: A Commentary on Hauser, Young, and Cushman, in *Moral Psychology, Volume 2: The Cognitive Science of Morality*, ed. by Walter Sinnott-Armstrong (Cambridge, MA: MIT Press), pp. 157–70.
- Rawls, John**, 1955. Two Concepts of Rules, *The Philosophical Review*, **62**: 3–32.
- Rawls, John**, 1971. *A Theory of Justice* (Cambridge, MA: Harvard University Press).
- Singer, Peter**, 1972. Famine, Affluence and Morality, *Philosophy and Public Affairs*, **1.3**: 229–43.
- Strawson, Peter F.**, 1974. *Freedom and Resentment and Other Essays* (London: Methuen).
- Turner, Piers Norris**, 2015. Punishment and Discretion in Mill's Utilitarianism, *Utilitas*, **27.2**: 165–78.
- Unger, Peter**, 1996. *Living High and Letting Die: Our Illusion of Innocence* (Oxford: Oxford University Press).
- Wolf, Susan**, 1992. Morality and Partiality, *Philosophical Perspectives*, **6**: 243–59.
- Zimmerman, Aaron**, 2013. Mikhail's Naturalized Moral Rationalism, *Jerusalem Review of Legal Studies*, **3**: 1–22.