The Macksey Journal

Volume 2 Article 124

2021

Moral Thinking and Common Morality: Integrating Two-Level Consequentialism and the Four-Principle Approach

Robert Campbell
University of Minnesota, Twin Cities

Recommended Citation
Campbell, Robert (2021) "Moral Thinking and Common Morality: Integrating Two-Level Consequentialism and the Four-Principle Approach," *The Macksey Journal*: Vol. 2, Article 124.

This Article is brought to you for free and open access by The Johns Hopkins University Macksey Journal. It has been accepted for inclusion in The Macksey Journal by an authorized editor of The Johns Hopkins University Macksey Journal.

Published by JHU Macksey Journal, 2021

Moral Thinking and Common Morality: Integrating Two Level Consequentialism and the Four-Principle Approach[1]

Robert Campbell

University of Minnesota, Twin Cities

Abstract

This paper defends a consequentialist framework integrating R.M. Hare's two-level consequentialism with Tom L. Beauchamp and James F. Childress's four-principle approach, which is grounded in common-morality theory. Hare argues consequentialism is most tenable when distinguishing between two levels of moral thinking: the critical level, a form of act consequentialism, and the intuitive level, a form of rule consequentialism. He argues that when the critical level selects its content, consequentialism justifies intuitive level thinking. I will argue Hare's framework is compatible with Beauchamp and Childress's four-principle approach because the principles satisfy the selection criteria Hare gives for intuitive, prima facie principles. After demonstrating how the four-principle approach is compatible with, complementary to, and justified by a two-level framework, I substantiate this integrative framework by discussing the four principles and four cases. I then respond to two objections, one by John-Stewart Gordon and another one I often receive. In doing so, I demonstrate the tenability and compelling properties of my integrative framework for consequentialists. I conclude by discussing limitations of this framework, future directions, and a summary.

Keywords: Two-Level Consequentialism, Utilitarianism, Four-Principle Approach, Principlism, Tom L. Beauchamp, James F. Childress, Richard Mervyn (R.M.) Hare, Normative Ethics

I: Moral Thinking and Two-Level Consequentialism

R.M. Hare was a 20th-century ethicist most recognized for his metaethical theory, universal prescriptivism. Here, however, we will discuss his work in normative ethics, which I will take as divorced from and tenable without his metaethics. Hare argues that act consequentialism, which demands we perform—from our available actions—the action that maximizes the good, and rule consequentialism, which demands we observe general rules that, when followed, maximize the good, necessarily collapse into one another. However, if rule consequentialism's rules are justified because their general observance promotes the good, then it is counterproductive that, in scenarios where we know with certainty that violating a rule will maximize the good, we must adhere to the rule regardless. However, if we accept exceptions to the rules and specify them to certain cases to maximize the good, we can factor in every relevant detail of the case and create hyper-specific rules, thus making rule consequentialism functionally indistinguishable from act consequentialism.[2]

As such, the terms act and rule consequentialism, Hare argues, are best used to distinguish between two types of moral thinking: critical and intuitive. The critical level is the version of act consequentialism outlined above, where rules—hereafter, principles—become hyper-specific. To exemplify how critical level thinking works, Hare asks us to imagine archangels "with superhuman powers of thought, superhuman knowledge and no human weaknesses."[3] These archangels incorporate all morally relevant details into their decision-making and construct a highly specific principle on which they act to maximize the good; hence, how rule and act consequentialism, perfectly applied, are functionally indistinguishable.[4]

However, we are imperfect beings so, to understand how we ought to think, we need to consider intuitive level thinking as well.

To exemplify how intuitive level thinking works, Hare asks us to imagine proles, "who have to rely on intuition and sound prima facie principles and good dispositions" all the time and are "totally incapable of critical [level] thinking."[5] Hare argues that sound prima facie principles are characterized by "firm and deep dispositions and feelings," which guide day-to-day ethical decision making and are used instead of critical level thinking in certain cases.[6] The prima facie principles on which these proles rely are analogous to rule consequentialism's rules; because proles can't think at the critical level, they can't make sound exceptions to their principles and construct the highly specific principles archangels do. However, if these prima facie principles are sound (i.e., justified by the critical level), they will generally promote the good because of how we select prima facie principles.

How we ought to think, on Hare's account, depends on our particular context and in what respects we resemble proles and archangels.[7] In certain cases, we have more time to think and better epistemic standing than in others, and we must adjust accordingly. Due to their nature, prima facie principles are best suited to cases with which an agent is familiar. Intuitive, prima facie principles, in such cases, eliminate the need to think at the critical level since the agent already knows what will likely best promote the good. However, if we have the epistemic standing to adequately understand a certain case and the time to consider which action would best promote the good, we ought to think at the critical level. Unfortunately, such cases are rare, due primarily to our limited capacities and the constraints upon us in realistic cases. Contrast such cases with ones where we must make a split-second decision. There isn't

enough time to employ the critical level, so we must act intuitively, relying on the feelings and dispositions engendered by our prima facie principles. However, these split-second decision-making situations are also rare. Many cases lie on a spectrum between these polarities, so the best approach to cases on it often lies between the intuitive and critical levels. In such cases, we might complement one level with the other by using the critical to parse and resolve conflicts between prima facie principles or to specify them to a certain case.

In sum, critical level thinking is best suited to complex scenarios when we have the necessary qualifications, and intuitive level thinking is best suited to familiar situations and split-second decision-making. However, for many cases, the best approach combines these complementary ways of thinking. Initially, this combination may seem unfeasible. Bernard Williams objected that we "cannot combine seeing the situation in that way, from the point of view of those [intuitive] dispositions, with seeing it in the archangel's way," where the intuitive level in and of itself is a means to maximizing the good.[8] This instrumentalization of the intuitive level also alienates the agent from it, thus undermining the value of acting intuitively and holding such dispositions.

There are two ways to diffuse this objection; the first is by appealing to a dual-process theory of moral psychology. Gary Varner thoroughly discusses this avenue. The theory claims that, empirically, we have an instinctual and immediate way of thinking, analogous to the intuitive level, and a slower and more deliberate way of thinking, analogous to the critical level, that we switch between depending on our context (e.g., our epistemic standing, time, etc.).[9] If a dual-process theory is true, it is psychologically feasible to combine the critical and intuitive levels since there is empirical evidence we do so already. However, empirical evidence doesn't

imply normativity, hence why another argument, articulated by Derek Parfit, is necessary. Consequentialism implies "we should cause ourselves to have, or to keep, the best possible sets of motives."[10] Thus, if a two-level framework is compelling and a dual-process theory of moral psychology is true, Williams's objection is diffused since we can, psychologically, use a two-level framework and have normative reasons to do so.

However, a critical question remains: which prima facie principles ought we to inculcate? Hare's answer is reminiscent of David Hume and Henry Sidgwick. He argues that many of our intuitive ethical responses and principles developed along consequentialist lines, such that they are, when relied upon in common situations, optimific.[11] This answer, however, is somewhat vague, and finding prima facie principles the critical level justifies can be a difficult task in and of itself. Although Hare outlines criteria for selecting prima facie principles and thoroughly discusses using them in a two-level framework, what exactly these principles are isn't fully explored. In what follows, I argue the four prima facie principles Tom L. Beauchamp and James F. Childress (hereafter, B&C) propose are suitable intuitive level principles. I will show how Hare's two-level framework is compatible with and justifies these principles, and how an integrative framework solves several problems with B&C's approach. Such an integrative framework, I argue, is tenable and compelling. To make these arguments, though, I first need to introduce the four-principle approach.

II: Common Morality and the Four-Principle Approach

B&C's four-principle approach—often called principlism—began as an approach to biomedical ethics but has since made some arguments (e.g., common-morality theory) that

make it applicable to normative ethics. The principles B&C propose are prima facie; however, they use the term differently than Hare. On their account, if an action is right or wrong prima facie, it is right or wrong "in the absence of other moral considerations" that would change our evaluation.[12] We will discuss these different definitions in section III. The first of the four prima facie principles B&C propose is respect for autonomy, which entails "respecting and supporting autonomous decisions;" the principle demands not only noninterference but promotions of autonomy.[13] For agents to be autonomous, B&C argue, they must meet three criteria: they must "act (1) intentionally, (2), with understanding, and (3) without controlling influences that determine their action."[14]

Respect for autonomy, however, may be overridden by another of the four principles (i.e., "other moral considerations"). For example, if person A autonomously attempts to murder person B, person C can justifiably override person A's autonomy by appealing to the principle of nonmaleficence, which entails "avoiding the causation of harm."[15] However, B&C don't argue the four-principle approach can't justify harmful action. Suppose the only way person C can prevent person A from murdering person B is by harming person A. In this case, person C's harm to person A is justified; although person C violates the principle of nonmaleficence, she does it to save person B. In this case, the principle of beneficence, which entails "relieving, lessening, or preventing harm and providing benefits and balancing benefits against risks and costs," overrides nonmaleficence.[16] What exactly beneficence entails depends on our particular context.

The context-sensitivity noted above applies to all these principles, including the principle of justice, which entails "fairly distributing benefits, risks, and costs."[17] This principle is

likely the most complicated one B&C propose. After discussing six theories of justice (viz., utilitarianism, libertarianism, communitarianism, egalitarianism, capability theories, and well-being theories), they conclude their normative discussion of the principle by writing that each theory of justice has insights that must be considered intelligently and argue that we ought to use these theories as resources when discussing certain cases.[18] This conclusion may seem unsatisfactory, but I will expand on it in section IV. To summarize, the four principles are prima facie, context-sensitive, and complementary; each is also justified by and grounded in common-morality theory.

B&C's common-morality theory claims the four principles are specifications of "universal norms shared by all persons committed to morality… [they are] applicable to all persons in all places, and we appropriately judge all human conduct by [their] standards."[19] Beauchamp argues these norms and the principles derived from them are proven by history "to ameliorate or counteract the tendency for the quality of people's lives to worsen or for social relationships to disintegrate."[20] Since history corroborates the efficacy of these norms as such, the principles derived from them have normative significance; hence, why moral people everywhere accept them.[21] However, common-morality theory accounts for cultural diversity by distinguishing between content-thin (i.e., formal) principles, which are universally accepted, and content-thick (i.e., particular) principles, whose acceptance is more contextual.

Formally defined, nonmaleficence is doing no harm, justice is treating equals equally, beneficence is doing good, and respect for autonomy is titularly defined. How do we do good, though? What constitutes equality? The formal principles are too content-thin to answer these substantive questions or be of any significant practical use; so, they are given content by

specifying them to certain cases, hence their context-sensitivity. B&C have focused on specifying them into a biomedical context, and we can also specify the principles to certain cultural contexts, hence their cultural sensitivity. As such, while moral people everywhere accept that we should be beneficent, just, nonmaleficent, and respectful of autonomy, what these formal principles practically entail depend on how they're specified. To clarify, by specification, B&C mean the "process of reducing the indeterminacy of abstract norms and generating rules with action guiding content."[22] As noted above, how we specify principles vary; moreover, these various specifications can lead to difficult moral disagreements, as Beauchamp notes.[23] To resolve such disagreements, we may make specific qualifications; but this is not B&C's only conflict resolution strategy.

Another strategy is balancing and weighing, where we consider the quality of the reasons we have to override a principle, the possibility of achieving the objective the overriding aims at, and whether preferable alternatives are available, the overridden principle is minimally violated, the violation's negative effects are minimized, and everyone is treated impartially.[24] In some cases (e.g., the example with persons A, B, and C above), which principles take priority are clear, but in other cases, it isn't. When balancing and weighing is ineffective, we may turn to the second strategy B&C endorse: the method of reflective equilibrium.

Reflective equilibrium, on B&C's account, builds off sound moral judgements made by individuals with the necessary epistemic and moral qualities to make a coherent moral framework. These individuals "must exhibit absence of prejudice, relevant knowledge, and honesty, as well as attitudes of sympathy and compassion for the welfare of others."[25] Reflective equilibrium, as such, works to make these individuals' sound moral judgements (e.g.,

specifications of the common morality's norms and four principles) into a coherent moral framework, adjusting their specifications and judgements when they conflict or seem to omit something important.[26] Practically, reflective equilibrium is an ongoing process, necessary to resolve conflicts between norms, principles, and specifications and improve particular moralities in doing so.[27]

However, I think these strategies are dubious. First, balancing and weighing is a sound process for moral decision-making, but its success depends on the reliability of the principles involved; the four principles are reliable, B&C argue, because they are justified by common-morality theory. Reflective equilibrium, on B&C's account, relies on qualified individuals whose epistemic and moral qualities are assumed to yield sound normative judgements with little argument but an implicit appeal to common-morality theory. Thus, the four-principle approach's tenability depends on common-morality theory. However, when B&C discuss three kinds of justification for common-morality theory—namely, empirical, conceptual, and normative—they don't follow through with any of them. As such, the four-principle approach is normatively ungrounded. This is a serious problem because, as Hare argues, although the intuitive level is necessary for standard agents, leaving prima facie principles ungrounded is a serious mistake since hard decisions and conflict resolutions between principles become arbitrary. In such cases and others where more critical thinking is needed, "To use intuition itself to answer such questions is a viciously circular procedure."[28]

However, in another, earlier article, Beauchamp does attempt to justify common-morality theory. He argues history proves common morality's norms and the principles derived from them promote "human flourishing by counteracting conditions that cause the quality of

people's lives to worsen," hence their normativity.[29] However, Beauchamp doesn't define flourishing nor delve further into this brief account; as such, even though he laid a groundwork, the four-principle approach remains normatively ungrounded. In the following section, I will justify the four principles as prima facie, intuitive level principles and show how a consequentialist justification of common-morality theory confers normative grounding to the four-principle approach and how other components of B&C's approach are compatible with a two-level framework, thus outlining our integrative framework.

III: Integrating Two-Level Consequentialism and the Four-Principle Approach

Let's begin by showing how an integrative framework defuses my critique of the four-principle approach in section II. If the four principles are justifiable intuitive level, prima facie principles in a two-level consequentialist framework, they are inherently lexical. The critical level would justify both their use and the methods that resolve conflicts between them. Moreover, the four-principle approach would add substantive content to the intuitive level. These arguments, however, rely on several assumptions. The first is B&C's four principles are justifiable prima facie principles; the second is B&C's conflict resolution strategies are compatible with consequentialism; and the third is consequentialism justifies and is compatible with common-morality theory. I will assume, for the following arguments discussing these assumptions, that consequentialism itself is tenable.

First, let's discuss the critical level reasons we have to accept beneficence, nonmaleficence, justice, and respect for autonomy as prima facie principles. Hare gives two basic criteria for prima facie principles: "the consequences of inculcating them in ourselves and

others," and "the size of the good and bad effects'' in the cases we would use these principles relative to "the probability or improbability of such cases occurring in our actual experience."[30] These criteria are rule consequentialist, meant to select principles the use of which, on the whole, will be optimific. The first criterion is important given Hare's definition of prima facie principles as deeply held commitments with effects on agents' moral psychology, and the second is important because it gauges the expected practical efficacy of a principle in consequentialist terms.

I argue the four principles meet these basic criteria, given the way B&C defend them. Recall common-morality theory. It claims history shows its principles effectively promote the good. If this is true, the consequentialist has reason to take them seriously. To help contextualize this argument, we need to consider Hume and Sidgwick. The former claims justice is normative since it is a necessary component of a functioning society, which has consequentialist value because, presumably, it facilitates a higher quality of life for its members; hence, it's in their common interest for people in a society "to regulate their conduct by certain rules."[31] Common-morality theory extends this claim not only to the principle of justice but to beneficence, respect for autonomy, and nonmaleficence.

As such, according to common-morality theory, when people accept these principles, they do it because history proves they facilitate a higher quality of life. The two-level consequentialist has reason to consider them as prima facie principles, then, because people are already primed to accept these principles, and they have been shown to promote the good. Now, consider Sidgwick, who observes how the common rules of a society differ from culture to culture. Depending on a society's environment and context, the rules regulating their behavior

will vary. However, these variations are grounded "in the real or perceived effects on general happiness of actions prescribed or forbidden by the code;" as such, a society's rules can develop to better promote the good according to its context.[32]

B&C mirror Sidgwick's claim, distinguishing—as noted earlier—between particular moralities and universal (i.e., common) morality. The principles of particular moralities typically build on the foundation of the common morality, specifying the formal principles like beneficence from simply doing good to doing good in a specific, culturally informed way.[33] In sum, to have a functioning society, as Beauchamp and Hume claim, people must accept the common morality's principles, and, as Sidgwick and B&C note, how people specify the principles depends, in part, on their environment, explaining both why people accept the common morality and particular moralities differ. Although I could do more to justify these arguments, taking them as basic premises for the following arguments is plausible; like Beauchamp, I don't think they claim too much.[34]

That said, these arguments alone don't suffice to show why these principles are normative or why they ought to be prima facie principles in a two-level, consequentialist framework; we must consider them in conjunction with Hare's criteria. Common-morality theory claims history shows the four principles to promote the good by facilitating society and the higher standard of living it produces. How a society specifies the four principles depends on its environment and context and varies according to what promotes the good best. Implicitly, a society's rules also vary according to which cases and contexts are regularly encountered in it. Thus, the particular morality of a society develops to promote the good of its citizens and is shown by history to be effective in doing so; if the common morality's principles were

ineffective, as Beauchamp notes, they wouldn't have developed.[35] Now, recall Hare's second criterion about the expected practical efficacy of a principle. Since history shows the common morality's principles generally promote the good and that they do so in relatively common cases, they meet this criterion.

The four principles meet Hare's first criterion for a connected reason. Namely, because history proves the four principles to promote the good, they're functionally normative, as Beauchamp claims.[36] Thus, the cost of inculcating these principles is small, meeting Hare's first criterion. This cost is especially small if, as Hume tenably claims, society educates and encourages its members to accept and internalize these principles because the society's survival is dependent on its members doing so.[37] If Hume is right, members of a society have both moral and nonmoral reasons to accept these principles; as such, the four principles meet both of Hare's criteria.

However, as B&C note, empirically demonstrating that people accept the four principles doesn't give the principles normative significance; further justification is necessary.[38] A two-level consequentialist framework provides such justification. Although the argument that the four principles developed to be integral to society doesn't itself demonstrate the principles are normative, according to a consequentialist justification, because the principles are necessary for society, which—we can assume—is good overall, the principles are valuable. Moreover, because they meet Hare's criteria, the principles are suitable prima facie principles, which are necessary for standard agents. Thus, we have sound normative reasons to accept and internalize these principles. By grounding common-morality theory and the four-principle approach in a consequentialist, two-level framework, common-morality theory and the four-

principle approach become normatively grounded, and the two-level framework's intuitive level becomes substantive; hence, how these approaches are complementary. Of course, there is more to the intuitive level and our integrative framework than this.

First, as I noted in section II, B&C use the term prima facie differently than Hare. While the former mean an evaluation independent of certain factors, the latter means a principle held deeply by an agent, with effects on her moral psychology. This leads to the question: are B&C's prima facie principles viable prima facie principles in Hare's sense? Yes, because of common-morality theory. As I noted in our brief discussion of moral education, people are already primed to accept and internalize the four principles; they have both a moral and nonmoral reason to in this case. The moral reason has to do with the normativity of these principles, and the nonmoral reasons have to do with the social consequences for violating or not accepting these principles. As such, it is difficult in many cases for agents not to internalize the four principles in the way Hare describes. Therefore, there is no difficulty in the difference in definitions here because, when we take common-morality theory to its logical conclusion, the definitions meaningfully intersect.

Now, let's consider another aspect of my critique in section II: conflicts between the principles. As discussed in section II, B&C have several conflict resolution strategies, like balancing and weighing, specification, and reflective equilibrium. The first two strategies are unambiguously compatible with two-level consequentialism. Balancing and weighing can be facilitated by the critical level, with the quality of our reasons for action being dependent on how well they promote the good; the remainder of the process is something consequentialists already do. As I noted in section II, balancing and weighing is a sound process, but its success

depends on the reliability of our principles. Since the four principles are justified by our integrative framework, they're reliable. Specification is compatible and something consequentialists already do, too. As Sidgwick observes, cultures specify their principles to promote the good, hence their consequentialist value. Moreover, as Hare argues, we should specify our principles to be optimific in certain cases (e.g., archangels' hyper-specific principles).[39]

Reflective equilibrium is more complicated for consequentialism. B&C argue the method of reflective equilibrium ought to begin with considered judgements that seem undeniable and from which qualified individuals build until they have a coherent set of moral beliefs. As I argued in section II, this strategy is dubious, given the numerous assumptions involved. However, because common-morality theory is now justified by consequentialism, B&C's implicit reliance on it is unproblematic. To make reflective equilibrium compatible with our integrative framework, then, I need only claim that taking a consequentialist norm as our considered judgement is as plausible as taking another norm as our basic, considered judgement.

If we grant that reflective equilibrium may begin by building off consequentialism, the method becomes integral to constructing a tenable two-level framework. At the intuitive level, reflective equilibrium would function as B&C outline, namely, by building out from the four principles to construct a coherent set of intuitive beliefs, dispositions, and principles. At the critical level, we would refine these phenomena, making the adjustments and qualifications needed for our integrative framework to be coherent and usable. As such, reflective equilibrium is compatible with and necessary for two-level consequentialism. Having clarified these aspects of our integrative framework, let's substantiate it.

IV: Substantiating our Integrative Framework

Now, let's consider—more substantively—how our integrative framework functions. Since consequentialism justifies common-morality theory, we should have another four, secondary considered judgments while using the method of reflective equilibrium: the principles of beneficence, nonmaleficence, justice, and respect for autonomy. What would this entail for each of the four principles? Specifically, how would the critical level refine them, and how would agents using our integrative framework think and act?

First, let's discuss beneficence and nonmaleficence. These are the principles most readily compatible with consequentialism. In a two-level framework, these principles would function, essentially, as intuitive versions of an agent-neutral, maximizing, consequentialist principle. Although the critical stereotype of act consequentialist agents as people who calculate the felicity of every action before they take it is often useless, its lesson is relevant here. Agents can't stop and consider every aspect of every case whenever they encounter it; they often don't have enough time or epistemic standing to effectively do so. We aren't archangels and acting as such—as noted in section I—wouldn't be recommended by consequentialism. As such, the prima facie principles of beneficence and nonmaleficence stand to prime agents to act intuitively, in the appropriate cases, as an act consequentialist would. Psychologically, this implies a hesitancy to do harm, a proclivity to do good, satisfaction when we act according to these principles, and dissatisfaction when we violate them. Importantly, this point about satisfaction applies to each of the principles.

Now, consider justice, which formally requires that we treat equals equally. As noted in section II, justice is perhaps the most complex principle B&C discuss. They don't outline or endorse a particular theory of justice but instead argue that each has lessons we should consider when making our decisions. As such, the form of justice at the intuitive level is also complex. Since our integrative framework is self-effacing, it can generally accommodate non-utilitarian theories of justice. Depending on our context, it might sometimes be better to be an egalitarian than a communitarian; we will discuss a case below to demonstrate how this works. While switching between these different theories of justice may sometimes be difficult, there is value in doing so, since each helps to enrich our analyses at the intuitive level, which we can then refine at the critical level. Moreover, if B&C can switch between these theories effectively in their four-principle approach, I see little reason why our integrative framework can't do the same, especially with its additional lexicality.

Lastly, consider respect for autonomy, the nuance of which is often overlooked by consequentialists. The most obvious argument for why consequentialists ought to respect autonomy is one of the best. Because hypothetically typical agents know what would best promote the good in their own lives and make ethical decisions, we ought not—in the absence of other moral considerations—to interfere with their autonomous decisions since they will, overall, be optimific. As such, we have sound, prima facie reasons to respect and promote autonomy; this implies a proclivity for truth-telling, non-interference, and an emphasis on informed consent. As with the other principles, though, these dispositions may be overridden by other moral considerations.

However, one difficulty arises from the occasional conflict between respect for autonomy and act consequentialism or beneficence, namely, paternalism. While in some cases the right course of action is clear, in others, it is more ambiguous. Paternalism is a multifaceted subject beyond the scope of this paper, but I see little reason why, if our integrative framework can adapt and justify the four-principle approach and its methods, it couldn't also adapt B&C's analysis of the topic and enhance it with the lexicality and attentive context-sensitivity of the critical level. Our integrative framework, then, provides a method—at least in theory—to resolve such conflicts. Although this method may sometimes be practically ineffective, and we may sometimes apply these principles imperfectly, this is unavoidable and unobjectionable. I will show how in section V.

Now, to demonstrate how this integrative framework might work, let's consider four cases. It will help to distinguish four categories of cases—distinguished by Varner—in which agents would use critical level thinking. The first is a novel case unfamiliar to the agent and to which, as such, intuitive level thinking is ill-suited; the second is when prima facie principles conflict; the third is an ongoing process, that of using the critical level to refine the intuitive level.[40] The last category is rarer than the other three: it is when something (e.g., an action) is intuitively wrong, but it is clear and sure that it would, in this case, promote the good better than the alternatives.[41] I will fill out these categories in the following paragraphs by discussing various cases and analyzing them through the lens of our integrative framework; in doing so, some important points will be clarified.

First, let's consider a case from B&C's specialty, medical ethics.

Medical Rationing: I'm a doctor with two patients; however, I can only treat one because of limited supplies. Patient A is a social hermit who has no friends or family; patient B is a social butterfly with many happy relationships. Patient B, therefore, has more social utility than patient A. Who should I treat?[42]

This case falls into Varner's second category, in which intuitive principles conflict; professional moralities also become relevant. As stated earlier, B&C specify the four-principle approach to a biomedical context; this is a particular, professional morality because agents in the biomedical profession are expected to adhere to its specifications. In this context, treating patients justly, on an egalitarian basis, is extremely important because medical professionals have established public trust with their professional specification of this principle. As such, although treating patient B because she has more social utility is beneficent, doing so violates medical professionals' specification of justice and would undermine trust in them. As such, in the long term, the benefits of treating patient B for her social utility would be outweighed by the expected negative consequences of harming the public's trust in medical professionals. Thus, in *Medical Rationing*, the critical level, facilitating balancing and weighing, recommends following the medical profession's particular morality.

Peter Singer made the second case we will discuss famous.

The Drowning Child: I'm walking past a shallow pond and notice a child drowning in it. There's no one else nearby, I could easily intervene, and, without my prompt help, the child will die. However, I will get my clothes dirty if I wade into the pond and help the child. Should I save the child?[43]

Clearly, I ought to save the child. But how would I think about this case if I adopted our integrative framework? This case doesn't fall into any of the four categories above since I can't effectively use the critical level because of time constraints. To do so would waste precious time and, because the critical level justifies the principle of beneficence, I know I can rely upon it when I make the intuitive decision to save the child. Moreover, I would have sympathy for the drowning child and the other mental and motivational phenomena that characterize a fitting agent in this situation, given the principle's psychological implications.

We may ask about the probability of this case and comparable ones in our actual experience, though, since this is one of Hare's criteria for prima facie principles. While cases like *The Drowning Child* are somewhat unrealistic and dramatized, they represent a category of cases that agents are likely to encounter. While the particular details are variable (e.g., the time agents have to think, the agents and patients involved, the costs and benefits of a beneficent action, etc.), there are many cases where an agent will be expected to act intuitively and beneficently. For instance, consider giving someone resources in a time of need, stopping to render aid, and other familiar situations. In such cases, if we act intuitively, as we are disposed to because of the principle, even if the case isn't time-sensitive, we will save more time for other things of value. The principle of beneficence, then, is well suited to *The Drowning Child*, despite its improbability. However, the same isn't true of other improbable cases.

Consider our third case, adapted from Parfit:

The Drowning Children: I'm walking past a shallow pond and notice four children drowning in it. Again, there's no one else nearby, I could easily intervene, and, without

> my prompt help, the children will die. I don't recognize three of the children, but the fourth is my child. The children are positioned such that if I save my child, I can't save the other three and vice versa. Should I save my child or the other three children?[44]

This case is tragic and in Varner's fourth category; not saving my child is intuitively wrong, but I'm sure that saving the other three children would be, at the critical level, better. While the case is improbable, it isn't impossible, and analyzing it will be informative. Let's first clarify the principles at play here; caring for our children is a specification of beneficence so widespread it can be readily conceived of as another principle entirely. This specification is widespread for good reason. As Parfit argues, with similar logic to Beauchamp's, caring for our children is—for what I take to be obvious reasons—an important principle for a functioning society, especially since such relationships, typically, are valuable for the people involved.[45]

As such, a society's members ought to—and likely will—accept this specification of beneficence, which implies strong love for one's children. Overall, this is good. However, in strange and improbable cases like *The Drowning Children*, this principle becomes problematic. Consequentialism implies it would be better if I saved not my child but the other three children. However, doing so, because of the strength of my love for my child, is motivationally improbable, if not impossible. So, suppose that acting intuitively, I save my child, and, because of my decision, the other three children die. What does our integrative framework imply? Much the same as what Parfit concludes: this is a case of blameless wrongdoing. Although I'm making the wrong decision, I should be blamed only to the extent beyond which my love for my child would be diminished, given *The Drowning Children's* improbability and the value of this

specification.[46] The same argument applies to other intuitive principles that have similar properties in comparable cases.

Lastly, consider a case adapted from Williams.

> *Coerced Killing*: I'm exploring the countryside when several soldiers approach and tell me they arrested thirty-four people for protesting the local government and that if I kill a prisoner, they will free the remaining thirty-three. If I don't kill a prisoner, the soldiers will murder every prisoner in their custody; should I kill a prisoner?[47]

This case is comparable, in some respects, to *The Drowning Children*, since—like caring for one's children—refraining from killing is a specification of nonmaleficence which is, for what I again take to be obvious reasons, important for a functioning society. However, how the principles manifest is substantively different.

Specified beneficence in the form of parental love permits few qualifications if the beneficent relationship is to survive. For example, suppose, in *The Drowning Children*, I somehow abandon my parental love and save the three other children. However, my initial judgement, that no one else could help, was mistaken, and a passerby saves my child. Talking to my child afterward, I try to explain my decision; while she might understand my reasons for acting as I did, my violation of parental love demonstrates that our relationship isn't as strong as it should be, thus damaging it and undermining the principle's purpose. By contrast, specified nonmaleficence in the form of refraining from killing allows more qualifications before the principle is damagingly undermined.

If I violate nonmaleficence by killing a prisoner in *Coerced Killing*, I can justify my decision to society in a way I can't with my child. While parental love has to do with specific agents, refraining from killing is a general obligation. Violating the former, whatever my reasons, damagingly undermines the relationship; violating the latter can be justifiable with sufficient reason since one violation of a general policy can be justified without damagingly undermining it. However, I may be blamed for killing a prisoner if we apply Parfit's conclusion in *The Drowning Children* to this case. Although I'm making the right decision, I should be blamed for it to the extent at which I still accept and internalize nonmaleficence because of *Coerced Killing's* improbability and the value this specification. Thus, the blame I receive may be minimal, but this is a case of blameworthy right-doing nevertheless. Again, we can make similar qualifications to other intuitive principles with similar properties in comparable cases. While there's benefit to discussing these and other cases further, for our purposes, this analysis is adequate. By responding to two objections in the following section, I aim to draw out further insights.

V: Objections

The first objection I'll discuss is one made by John-Stewart Gordon, who—in an article defending the four-principle approach—argues that

> Utility and consequences are only part of the moral discourse in principlism; motives and actions, characters and virtues, and the prima facie universal principles—justice, autonomy, beneficence, and nonmaleficence—are other important aspects that enrich the moral discourse in general and that allow it to persuasively address cross-cultural

> moral problems in particular. The hypothesis that all (cross-cultural) moral problems can be solved by the application of the notion of utility seems somewhat awkward.[48]

This objection, however, doesn't account for more sophisticated versions of consequentialism than Benthamic utilitarianism.

Repeatedly, philosophers have demonstrated consequentialism's ability to accommodate and justify non-consequentialist concepts. Parfit demonstrated the tenability of self-effacing consequentialism by showing that, although consequentialism can be indirectly self-defeating, this is theoretically unproblematic because consequentialism can't be directly self-defeating.[49] Further, self-effacement isn't a potent objection to consequentialism because, even if consequentialism was wholly self-effacing, this would have little to do with the normative truth or falsity of consequentialism. [50] Parfit's arguments are compatible with our integrative framework especially, given his remark that consequentialism may demand that agents don't have only certain "desires and dispositions, but corresponding moral emotions and beliefs" as well, which parallel how prima facie principles function in our integrative framework.[51] Moreover, if the concepts Gordon mentions are as useful as he claims, consequentialists have reason to incorporate them into their frameworks, like I did with B&C's conflict resolution strategies in section III.

With this common objection addressed, let's move onto the second, which claims that

> since we are not archangels, we can't adjudicate between the principles effectively all the time or even know if the common morality's principles are the most effective. Therefore, we are not able to use your integrative framework as you claim.[52]

This objection's premise is true; however, the conclusion doesn't follow. First, not being able to resolve conflicts between the principles at all times is characteristic of morality; there are some moral questions to which, practically, we don't know the answer. If we were archangels, we would be able to answer every moral question. But we aren't archangels. This doesn't imply, however, that we can't employ critical level thinking to great effect. As the cases in section IV demonstrate, even with our limited capacities, the critical level is useful. As Hare notes, the question here is a practical one. Its answer "depends on what powers of thought and character each one of us, for the time being, possesses."[53] While this implies some subjective judgement, such subjectivity seems ineliminable in any tenable moral theory.

That said, although we can't use the critical level perfectly, we can at least use it to find and justify prima facie principles, resolve conflict between them, and approximate critical level thinking under the right circumstances. When we take the time to think through such principles—as I have done here—our thought processes are less subject to error. Considering the empirical and normative components of these principles, which have been adequately discussed by now, the justification of them in section III is satisfactory for now. However, there are some issues in specifying and making these principles coherent. As noted in sections II and III, intercultural and interpersonal disputes can occur even if people agree on consequentialism and common-morality theory. While such disagreements can lead to dialectical improvements in particular moralities, they can lead to difficult moral disagreements as well. Grounding these principles in consequentialism—assuming consequentialism is true—will help resolve these disagreements. Although such conflict resolutions will be difficult, they will be more productive than simply going in circles, ceaselessly arguing, and making little progress because we have no

sound method to resolve our disagreements. Practical ethics is difficult, but the critical level of our integrative framework provides a way to think through these difficulties soundly, and the intuitive level accounts for our realistic constraints.

VI: Neglected Points, Future Directions, and Summary

Now, let's discuss some points I neglected and future directions. First, I purposefully stating what constitutes "the good" we are bound by consequentialism to promote. Value theory is beyond the scope of this paper, so I left it an open question; with some adjustments, I believe our integrative framework is compatible with several competitive value theories. Additionally, although I briefly discussed moral psychology in several sections, I neglected its empirical components because I couldn't give them adequate attention in this paper. As such, normative theorizing will have to do; Varner discusses moral psychology in greater depth, however. I also assume consequentialism is a tenable moral theory because justifying it, here, is impossible in this paper.

A subject of future interest is instrumentalization and alienation. Our discussion of this topic in section I is adequate but, given the topic's complexity, another project discussing it is necessary to do it justice and contribute something more interesting. Similarly, further discussion of moral blame—especially its public aspect—and the psychological implications of Hare's prima facie principles is warranted. Another point of interest is the virtues and what Douglas W. Portmore calls reason-responsive attitudes. The way Hare defines prima facie principles has some similarities with these phenomena, and parsing the differences between these concepts and judging their compatibility with a two-level framework seems worthwhile.

Finally, further discussion of the cultural component of common-morality theory and the relative conservatism of our integrative approach is fascinating and necessary.

To summarize, Hare's two-level framework is tenable independently of his metaethics and enhanced by integrating it with B&C's four-principle approach, which—despite its practical focus—makes important normative claims. By grounding common-morality theory in consequentialism, it becomes tenable, the four-principle approach becomes normatively grounded, and our integrative framework is substantiated by it. Despite objections, our framework is practically tenable, as I demonstrate by discussing four cases. In this paper, my aim isn't to argue our integrative framework is the best available but to demonstrate its tenability and compelling properties for consequentialists. Its assumptions and further implications are worth exploring.

End Notes

[1] Sincere thanks to Michelle Mason Bizri, Tamara Fakhoury, Sarah Holtman, Kiley R. Komro, and Kylie Shahar for their feedback, and audiences at the Washington and Lee University's Mudd Center 2021 Undergraduate Ethics Conference and Johns Hopkins University's 2021 Macksey Symposium for their questions and comments on my integrative framework. Thanks also to the staff at *The Macksey Journal*, especially Daniel Trovato for their guidance throughout the publication process.

[2] Hare, *Moral Thinking*, 43; n.b.: although Hare uses the term utilitarianism to describe his normative framework, I will use the term consequentialism instead, since this paper doesn't take any stances on value theory.

[3] Hare, *Moral Thinking*, 44.

[4] Hare, *Moral Thinking*, 44.

[5] Hare, *Moral Thinking*, 45.

[6] Hare, *Moral Thinking*, 38-39.

[7] Hare, *Moral Thinking*, 45.

[8] Williams, "The Structure of Hare's Theory," 189-190.

[9] Varner, *Personhood, Ethics, and Animal Cognition*, 84-88.

[10] Parfit, *Reasons and Persons*, 49.

[11] Hare makes this claim several times (e.g., *Moral Thinking*, 48-49 and 63-64). Hume and Sidgwick are cited below.

[12] Beauchamp and Childress, *Principles of Biomedical Ethics*, 15.

[13] Beauchamp and Childress, *Principles of Biomedical Ethics*, 102-103.

[14] Beauchamp and Childress, *Principles of Biomedical Ethics*, 102.

[15] Beauchamp and Childress, *Principles of Biomedical Ethics*, 13.

[16] Beauchamp and Childress, *Principles of Biomedical Ethics*, 13.

[17] Beauchamp and Childress, *Principles of Biomedical Ethics*, 13.

[18] Beauchamp and Childress, *Principles of Biomedical Ethics*, 281.

[19] Beauchamp and Childress, *Principles of Biomedical Ethics*, 13. Since common-morality theory is also defended in one of Beauchamp's earlier articles, I will reference it as well.

[20] Beauchamp, "A Defense of the Common Morality," 261. Italicization omitted.

[21] Beauchamp, "A Defense of the Common Morality," 261.

[22] Beauchamp and Childress, *Principles of Biomedical Ethics*, 17.

[23] Beauchamp, "A Defense of the Common Morality," 268.

[24] Beauchamp and Childress, *Principles of Biomedical Ethics*, 23.

[25] Beauchamp and Childress, *Principles of Biomedical Ethics*, 440.

[26] Beauchamp and Childress, *Principles of Biomedical Ethics*, 441.

[27] Beauchamp and Childress, *Principles of Biomedical Ethics*, 442.

[28] Hare, *Moral Thinking*, 40.

[29] Beauchamp, "A Defense of the Common Morality," 260.

[30] Hare, *Moral Thinking*, 40.

[31] Hume, *A Treatise of Human Nature*, 377-379.

[32] Sidgwick, *The Methods of Ethics*, 455.

[33] Beauchamp and Childress, *Principles of Biomedical Ethics*, 5.

[34] Beauchamp, "A Defense of the Common Morality," 261.

[35] Beauchamp, “A Defense of the Common Morality,” 261.

[36] Beauchamp, “A Defense of the Common Morality,” 261.

[37] Hume, *A Treatise of Human Nature*, 375.

[38] Beauchamp and Childress, *Principles of Biomedical Ethics*, 452.

[39] Hare, *Moral Thinking*, 51.

[40] Varner, *Personhood, Ethics, and Animal Cognition*, 15.

[41] Varner, *Personhood, Ethics, and Animal Cognition*, 16.

[42] This case and my response draw from my unpublished paper, “Medical Rationing, Moral Thinking, and Common Morality.”

[43] Singer, “Famine, Affluence, and Morality,” 431.

[44] Parfit, *Reasons and Persons*, 33.

[45] Parfit, *Reasons and Persons*, 32.

[46] Parfit, *Reasons and Persons*, 33.

[47] Williams, “A Critique of Utilitarianism,” 98.

[48] John-Stewart Gordon, “Global Ethics and Principlism,” 269.

[49] Parfit, *Reasons and Persons*, 53-55.

[50] Parfit, *Reasons and Persons*, 43.

[51] Parfit, *Reasons and Persons*, 40.

[52] Thanks to Michelle Mason Bizri and others for posing this objection.

[53] Hare, *Moral Thinking*, 45.

Bibliography

Beauchamp, Tom L. “A Defense of the Common Morality.” *Kennedy Institute of Ethics Journal* 13, no. 3 (September 2003): 260-261. https://doi.org/10.1353/ken.2003.0019.

Beauchamp, Tom L., and James F. Childress. *Principles of Biomedical Ethics*. Eighth edition. New York: Oxford University Press, 2019.

Gordon, John-Stewart. “Global Ethics and Principlism.” *Kennedy Institute of Ethics Journal* 21, no. 3 (September 2011): 251-276. http://doi.org/10.1353/ken.2011.0011.

Hare, R.M. *Moral Thinking: Its Levels, Method, and Point*. New York: Oxford University Press, 1981.

Hume, David. *A Treatise of Human Nature*. New York: Barnes and Noble, Inc., 2005.

Parfit, Derek. *Reasons and Persons*. Reprint. New York: Oxford University Press, 1987.

Sidgwick, Henry. *The Methods of Ethics*. Seventh edition. Indianapolis: Hackett Publishing Company, 1981.

Singer, Peter. “Famine, Affluence, and Morality.” *Philosophy & Public Affairs* 1, no. 3 (1972): 229-43. http://www.jstor.org/stable/2265052.

Varner, Gary E. *Personhood, Ethics, and Animal Cognition: Situating Animals in Hare’s Two-Level Utilitarianism*. New York: Oxford University Press, 2012.

Williams, Bernard. “A Critique of Utilitarianism.” In *Utilitarianism: For and Against*, edited by J.J.C Smart and Bernard Williams, 75-150. New York: Cambridge University Press, 1973.

Williams, Bernard. “The Structure of Hare’s Theory.” *In Hare and Critics: Essays on Moral Thinking*, edited by Douglas Seanor and N. Fotion, 185-196. New York: Oxford University Press, 1988.