ABSTRACT

Title of Dissertation:	A THEORY OF LEADERSHIP AND ITS APPLICATIONS
	Leisa Elizabeth Schwab, Doctor of Philosophy, 2023
Dissertation directed by:	Professor John F. Horty, Philosophy

No system of laws and political institutions is without gaps, and leaders are required—often in the face of uncertainty and under a heavy burden of risk—to fill them. This project adopts a view of individual leadership that finds its roots in the ancient world with Plato, but which speaks to modern problems like the role of appointed administrative officials in a complex democracy and the problems of autonomous weapons. It is composed of a series of papers exploring this gap-filling leadership activity in a modern democratic state from both normative and descriptive perspectives. The first paper, "Making Ourselves Accountable: An Ethics for the Administrative State" addresses the discretionary decision making by un-elected officials through which many of our society's important leadership decisions are made. It argues for the necessity of these leaders and recommends criteria to guide their decision making in conformity with contemporary democratic ideals. The second paper, "Seeking Standards for Leadership Reasoning in the Executive Branch by Analogy to Representation and Judicial Reasoning," looks deeper into the work of such leaders to better understand the place of their role in shaping the law alongside legislative representation and judicial discretion. The third paper, "A Different Kind of Responsibility Gap: Trust and the Burden of Risk as a Limit on Military Automation" considers the problem of autonomous weapons in the

context of this theory of the individual leader as a necessary component within the legal and institutional system. Inspired by ancient notions of the activity of governing as an activity fundamentally about leaders before it is about laws, it argues that even fallible human leaders who fall short of the ideal remain necessary no matter how sophisticated or accurate an automated system we may devise.

A THEORY OF LEADERSHIP AND ITS APPLICATIONS

by

Leisa Elizabeth Schwab

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Professor John F. Horty, Chair
Fabrizio Cariani
Brian Kogelmann
Rachel Singpurwalla
Karol Soltan

Table of Contents

List of Figures

Introduction: The Place of Leadership in Democracy: Modern Problems, Ancient Roots

Until philosophers rule as kings in cities or those who are now called kings and leading men genuinely and adequately philosophize, that is, until political power and philosophy entirely coincide…cities will have no rest from evils, Glaucon, nor, I think, will the human race. –Socrates, Plato's Republic

The requirement…if they are going to produce a good imitation of that true constitution of one man ruling with expertise, so far as they can, is that—given they have their laws—they must never do anything contrary to what is written… –The Visitor, Plato's Statesman

Is it a god, my friends, who in your view should take the credit for your legal arrangements? Or some human? –The Athenian, Plato's Laws

If we were asked to explain what a government is, we would likely say something about a structure of laws and institutions that wields authority in the name of a political state. This is strange because the activity of governance that a government carries out is not first and foremost a business of either laws or institutions. The business of government is a human business, accomplished primarily not by laws but by human beings. Human beings are fallible, changeable, and unpredictable. No individual leader could hope to personally resolve every disagreement and dilemma facing the society she leads. We establish laws to meet as many of those needs as we can foresee in the best and most consistent ways we can, but that is not where the story ends. There will always be gaps in the law because no code of law can address every possible need of the people living under it, gaps that must be filled dynamically by a human leader. This interdependence between laws and human leaders carves out the sometimes puzzling place of individual leaders, and leader discretion, in modern democratic governments. We have them because we require them, and that requirement is nothing new.

The *kallipolis* of Plato's *Republic*, we are told, is the happiest of all possible cities because it is ruled by a philosopher king who is wise and possesses true knowledge (Plato, trans. Grube 1992, 148-149). Such a person would be best qualified to create truly good laws and to lead citizens in following them (473d-e; 158, 484b-d). The *Statesman* returns to the idea that the best possible society

is one ruled by a truly knowledgeable leader but says more about what this leader does and how they might fit into a structure of laws and institutions (Plato, trans. Rowe 1999). This true statesman would be the one who could provide the best possible set of written laws to society, but their expertise would not be expertise only about the law and their true work is not the work of legislation alone (69, 295e-296a). It is expertise about the nature and needs of the people that they lead. The true leader knows what is best and most just because they understand the needs and natures of their people (71, 297b). They are able to make society strong and harmonious for the same reason (95, 310e-311).

Written laws, we are told, are only a poor substitute for the governing work of the true leader (71, 297d). No law could ever "accurately embrace what is best and most just for all at the same time…" because "practically nothing in human affairs ever remains stable…" (66, 294b). There will be gaps in even the best laws, made by the best leader because they are only general prescriptions of what is most correct for most people in most situations. (68, 295a). They may bend themselves "more or less towards" "what is best and most just," but they cannot reach it (66, 294a-c). We require them anyway. Even the expert leader cannot personally oversee every decision in the society they lead (67-68, 295a-b), and if we lack the true expert leader to guide us, we must settle for strict adherence to a code of laws as the best available substitute (71, 297d-e).

In the *Laws*, the theme of laws, lawmaking, and governing being in some fundamental way about understanding and addressing the individual needs of the people as best as possible remains. In that dialogue, the interlocutors set up a system of laws for an actual city, *Magnesia*, instead of the "best possible city," which as described closely resembles the *kallipolis*. "A city of that kind…," one of the interlocutors says, "I don't know if its inhabitants are gods or a number of sons of gods" (Plato, trans. Griffith 2016, 186, 739e). *Magnesia* will be a city not for gods, but for men. It will be

ruled under a detailed code of laws, but as always, the written proscriptions of those laws alone can't quite capture the full work of governing the city:

> the lawgiver's task, in all probability, goes beyond merely putting laws into place and calling it a day; something more is called for, over and above the laws, something whose nature puts it midway between advice and actual law (289, 822d-e).

Laws alone cannot accomplish that work, because any city or society is fundamentally an association of people, and any association requires a leader (52, 640-641).

In our own contemporary Magnesia, we have detailed constitutions that establish networks of institutions and codes of law and regulation so detailed they require a library—or powerful computer system—to record them. We also have legions of human leaders given authority to fill the endless gaps within that system. Our "city" is not a city of gods, and these leaders may not be in all areas true expert statesman. The *Statesman's* admonition that such a society should avoid error by refusing to ever do anything "contrary to what is written" in the law cannot guide us where the law is silent, or when it offers only "something whose nature puts it midway between advice and actual law" (1999, 71-72, 297d-e; 2016, 289, 822d-e). How, then, can they best perform the work of leadership in such cases?

This project adopts a view of the nature and limitations of law, and its relationship to human leaders, that owes its roots to Plato, and the shift in the dialogues from discussions of an ideal society led by the Philosopher King to talk of what all other societies must do to reach as close to that ideal as they can in such a leader's absence. It is a concept of leadership that remains rooted in the aspiration towards wisdom and truth as ideals for decision making, but which lives in the non-ideal world of imperfect leaders working to shape incomplete laws around the bends and spaces of diverse and ever-changing human needs as much in the direction of what is "best and most just" as it is possible for them to go in an unavoidably messy world. That world is messy largely because it is full of human beings, and human beings are messy. They are dynamic, unpredictable, and often

irrational. Leadership within government always places the leader in a relationship with both the laws and the people they must lead.

The leadership called for in these moments is not the motivational model of a leader whose task is primarily to motivate others to pursue a certain course of action or to share some common goal, though this type of leadership decision making is a necessary step in that process since someone must first select that course or identify that goal. This leader is an individual decision-maker, the person who must find or make the way forward when the way is not already clear and identify which option of a uniformly bad bunch is nearest to some unattainable best. Such a leader may be a high-profile public figure whom we would naturally and easily identify as such, and likewise they may then proceed to engage in the familiar steps of motivational or commitment-based leadership—but they may not be. This type of leadership, which we might call individual decisive leadership, is centered around the responsibility and authority to decide. It manifests in the moment of decision making. It does have a relational aspect, in that the leader's decision puts her into a relationship with a member or segment of the public that her decision will affect where she is, for purposes of that decision, the embodiment of important government authority. However, in cases of motivational or commitment-based leadership, both the activity of that leadership and its desired result center around a leader's relationship with those she leads—her actually *being a leader* both requires and establishes a certain relationship with them. In cases of individual decisive leadership, on the other hand, though they take place in the context of a certain relationship between the decision-maker and those affected by the decision which shapes norms for the process, the leadership activity that takes place is a decision-making activity, not a relationship-building or relationship-leveraging activity. It is the responsibility the decision-maker bears towards those who will be affected by her decision and those who have entrusted her to make it that makes the relationship one of leadership—at least for the duration of that decision. She may then go on to

engage in more familiar forms of leadership, or she may return to bureaucratic obscurity until the next such decision is required.

I have explored this activity in a series of papers considering what it means, in a complex contemporary government, for these leaders who are not "philosopher kings" to nevertheless do their best to "genuinely and adequately philosophize" (1992, 148, 473c-d) and why, thousands of years after the dialogues were written, we still need them to complement our systems of written law in much the same ways, and for the same reasons, as those cited by Plato. The project is shaped by my background as both a military officer and military attorney practicing operational law and as a civilian government attorney for an administrative agency, as is the non-standard concept of leadership that unites it.

The first paper, "Making Ourselves Accountable: An Ethics for the Administrative State" addresses the discretionary decision making by un-elected officials through which many of our society's important leadership decisions are made. It argues for the necessity of these leaders to supply reasons in otherwise rationally underdetermined cases and recommends criteria to guide their decision making in conformity to our contemporary democratic ideals for what is good and just in government. In that context, the activity in question rises above mere bureaucracy to the level of leadership by virtue of the discretionary judgment required to select and then prioritize reasons for decision in cases of public trust and importance and the burden of personal responsibility that accompanies such authority. The second paper, "Seeking Standards for Leadership Reasoning in the Executive Branch by Analogy to Representation and Judicial Reasoning," looks deeper into the work of such leaders to better understand the place of their role in shaping the law alongside legislative representation and judicial discretion. It pays particular attention to the way that epistemic limitations and their associated risks shape decisive processes in the executive branch, and the way that those burdens of responsibility again place those officials into a leadership role at the moment

of decision, even if only for that moment. The third paper, "A Different Kind of Responsibility Gap: Trust and the Burden of Risk as a Limit on Military Automation" considers the problem of autonomous weapons in the context of a theory of the individual leader as a necessary component within the legal and institutional system. Inspired by ancient notions of the activity of governing as an activity fundamentally about leaders before it is about laws, it argues that even fallible human leaders who fall short of the rational ideal remain necessary no matter how sophisticated or accurate an automated system we may devise to act in their place. Governing is always a relational activity among human beings. The health of that relationship requires a measure of trust on the part of those governed, and that trust requires in turn trustworthiness on the part of those doing the governing that cannot be replicated by machines.

Works Cited

Plato. *The Laws*, edited by Malcolm Schofield, translated by Tom Griffith, Cambridge University Press, Cambridge, 2016. Cambridge Texts in the History of Political Thought.

--. *Republic*, translated by G.M.A. Grube, revised by C.D.C. Reeve, Hackett Publishing Company, Indianapolis, 1992.

--. *Statesman*, translated by Christopher J. Rowe, Hackett Publishing Company, Indianapolis, 1999.

Paper 1: Making Ourselves Accountable: An Ethics for the Administrative State

On January 14th, 2022, the United States Federal Register[1] reported a total of 112 filings *for that day alone*, to include 85 notices of various administrative body meetings, 8 proposed agency rules, and 19 (finalized) administrative rules—a rather average day of publicly reportable activity in the U.S. Administrative State. There are over 430 departments, agencies, and sub-agencies (such as bureaus) within the U.S. Federal Government, split into various headquarters and subordinate units, with offices scattered throughout the nation. Collectively, the Federal Government employees over *four million* people, a significant share of whom are empowered to exercise discretion on behalf of the government in at least some limited area in order to discharge their daily duties. (Congressional Research Service, 2021, 6). The authority exercised by such officials is the most direct and pervasive vehicle through which individual citizens encounter government authority in their daily lives. These are the officials responsible for making recommendations for how long we ought to remain isolated following a positive COVID-19 test, whether our income tax filing require a second look, or whether we should be granted a lease to graze livestock on federal land. They even include the administrative law judges who are charged with determining whether all of these other decisions have violated our rights by being made in a way that is arbitrary and capricious—the limited avenue of recourse available in most such matters.

These administrative officials must make difficult and important decisions as leaders within the government in situations where legal and regulatory authorities are silent or vague, leaving outcomes up to individual discretion. The problem is one that characterizes contemporary governance, but it is not itself a modern problem. These sorts of gaps in the business of governing are a manifestation of the same sort of uncertainty Plato acknowledges in the *Statesman,* that no

[1] federalregister.gov

system of written laws could possibly account for every situation that will arise and decision which will have to be made in the activity of ruling over people and so individual leaders—hopefully exhibiting the appropriate expertise—are also required. (1999).[2] What is needed, though, is not an understanding of what *one* "true Statesman," or truly competent, ethical, or effective leader must do—but many thousands of them.

This paper will examine how inescapable vagueness in the legal and regulatory system necessitates the discretionary authority these leaders wield, and examine three contemporary theories of democratic authority and leadership, and the proper position of administrative branch authority within a democratic system. I will explain in turn how each of these three theories offers useful insights to the problem at hand and where they fall short.

In particular, each of these theories struggles to deal with three classes of hard cases that frequently confront administrative leaders. The first type are cases involving the sort of specialized epistemic demands which many expert administrative agencies were specifically created to handle, as well as cases that are epistemically opaque for various practical reasons: cases where Nobody Knows what ought objectively to be done. The second type are cases of great complexity, obscurity, or (whether in absolute terms, or absent aggregation) low stakes: cases where Nobody Cares. The third are cases where fundamentally irreconcilable interests are in competition with one another, and/or where all available courses of action are publicly unpopular or otherwise unappealing, yet a decision must still be made: cases where Nobody Can Agree. I will propose that officials facing these sorts of cases should proceed by making themselves accountable in a normatively significant way, by

[2] In that dialogue, Plato's interlocutors characterize a system of written laws as a sort of "second-best" option when a Monarchy with a wise ruler (like the philosopher king of the *Republic*) is unattainable. There's some brief acknowledgement that these systems have limitations in dealing with problems that were not anticipated at the time the laws were created, and so some sort of "king" or other leader also may still be called upon. The "true" statesman is the expert ruler who engages in the "art of kingship," which is a "weaving together" of the different dispositions of the citizens--courageous, cautious, and vicious—to accomplish those parts of rulership that a set of written laws "chasing after the truest constitution" still cannot encompass. (1999, 78, 94-96.) Much more is said about what these laws themselves might be in the *Laws*.

adhering to three principles in their decision making: Assumption of Risk, Dedication to the Public and the Institution, and Identification and Ordering of Priorities. A non-elected leader possessing discretionary government authority who does so will be able to resolve otherwise underdetermined cases like those in my three problem classes in a way that is neither arbitrary nor capricious, and consistent with democratic values.

Section I: Vagueness and Discretion in the Routine Business of Governance

This section will present an example of how and why a major portion of leader decision making in the United States takes place within the vast administrative state, encompassing everything from land management, to intelligence, to the Departments of the military. These decisions are made outside of the direct flow of voter choice and candidate persuasion or representative accountability through which we might typically be inclined to analyze the exercise of democratic power. Discretionary decision making of this sort is a required feature of any effort to integrate a comprehensive rule of law with a network of enduring political institutions. The reason for this has to do with the relationship between vagueness and discretion in any system of laws, and the influence of uncertainty on their creation, administration, and interpretation.[3]

[3] For the purposes of this discussion, I am concerned with the effect vagueness in the law has for placing government authority into the hands of numerous un-elected executive-branch officials, with the claim that this is both acceptable and necessary for a well-functioning democratic republic, and with the resulting questions surrounding how the use of such power ought to be guided. As a general matter, thought, the ideas both that vagueness is an unavoidable aspect of the law, and that it does useful work in the legislative and interpretive processes, is well discussed by both philosophers and legal theorists – though they may disagree about exactly what role vagueness plays in either design or application, and to what extent its effects should be seen as somehow intentional. See, for example, Waldron (1994) (arguing that, despite the "void for vagueness" legal doctrine, there are instances of vagueness in the law that have instrumental value, and which do not constitute flaws); Asgeirsson (2015) (a logic-focused piece arguing that thinkers like Waldron are mistaken in attributing instrumental value to vagueness itself, when instead the useful work is done by a semantic phenomenon of which vagueness per-se is only a part, for somewhat technical reasons—but not disagreeing that these phenomenon play a role in providing necessary flexibility that lawmakers are unable to include in the formulation of black-letter law); D'Amato's (2011) (construing vagueness as unavoidable, and as importantly prompting judges to interpret laws in accordance with notions of justice, ultimately leading to better outcomes), and, of course, Hart's (1958), famously discussing the idea that the work of judicial interpretation consists of resolving indeterminacy left by unavoidable vagueness in the language of laws, in what he calls "penumbra" cases.

Most occasions requiring government decision makers to exercise their individual discretion as leaders at various levels arise from vagueness in law and policy. This vagueness proves to be inescapable for several reasons. These include practical limitations on a legislature or administrative body's ability to anticipate all future decisive points and relevant factors, both making it impossible to create guidelines addressing all eventualities (so that in some cases the intent of the legislature will be inadvertently unclear), and requiring that law or rule structures be put in place with enough vagueness to allow for flexibility in unforeseen circumstances (the legislature having remained silent by design). These two factors are intertwined at various levels of governance, with vagueness in laws and rules giving rise to the uncertainties that require the exercise of discretion, and the need for discretion in the face of uncertainty about future decisive points in turn requiring a certain level of facilitating vagueness in laws and rules by design.

This point about the respective roles of vagueness, uncertainty, and discretion in the law is itself neither novel, nor particularly controversial. On the back end, the after-the-fact of government activity, Hart describes this inescapable vagueness in law as the "penumbra" of uncertainty surrounding all legal rules, where a judge—as an individual empowered with the discretion to resolve such uncertain cases—must determine how to apply general rules to a particular case because "[f]act situations do not await us neatly labeled," and "someone must take the responsibility of deciding" how the rules are to be applied. (1958, 607). Beerbohm's portrayal of the intention-setting function of democratic leadership is similar, drawing on Raz's description of individual decision making in cases where no one course of action is more rational than another to understand such vacuums of rational determination ("an underdetermination of reasons" (2015, 643)).

On the front end, the in-the-moment exercise of government authority, most of this uncertainty exists, and most discretion is exercised, behind the scenes of the political debate stage in day-to-day administration by appointed officials, outside of the awareness-and interest-of the great

majority of the electorate. There are several ways this comes to be the case. Sometimes it is because each of the competing interests involved present reasons for decision in their favor, so that "the electorate," or at least some segment of it, has expressed a preference but popular opinion is nonetheless uncertain, so that models of democratic authority as properly following majority opinion fail to indicate the correct course of action. Sometimes, it is because the decisions to be made are either technically complex (and obscure) enough that popular opinion would be a poor guide in the matter even if it were clearly expressed (so that majority opinion fails us), but also uncertain in the view of technical experts (so that the empowerment of certain administrative officials due to their superior epistemic qualifications in their area of discretion also provide no guidance). Often, it is because there simply *is* no clear epistemic basis for prioritizing one course over another, all courses will have some undesirable consequences, and the matter to be decided is of little interest to the electorate at large (though some segments of it may be intensely interested).

Mundane though these decisions may be, however, they are anything but simple below the surface, both in terms of impact, and in terms of the complexity of their position within the structure of U.S. Federal Government power. In the U.S. this sort of extensive reliance on numerous mid- and low- level individual leaders to exercise discretion in areas of vagueness is amplified at two levels by the structure of Federal legislative and rulemaking activity, and the relationship between elected bodies and executive institutions, that are required to affect the governance of such a large and diverse nation with all its citizens' competing needs and interests.

Ongoing legal disputes about the Bureau of Land Management (BLM) Wild Horse and Burro Program, and its need to nevertheless conduct day-to-day operations in accordance with its statutory mandate, offer a comprehensive example of how so much vagueness ends up in the laws parceling democratic authority out to non-elected entities, and why there is so much need for discretion in the daily business of governing under those grants of authority.

Thousands of wild[4] horses[5] live on public lands in the United States—most of them in areas of the West where that land is also in demand for cattle grazing. The needs of free-roaming animals conflict with the interests of cattle ranchers (primarily) and other users of public lands—and the land can only sustain so many animals. (Davies and Boyd, 2019). In the mid 20th century, wild horses on Western rangelands were routinely rounded up for commercial purposes or shot by ranchers to eliminate their interference with cattle grazing.

By 1971, with wild horses "fast disappearing from the American scene," Congress passed the Free-Roaming Wild Horse and Burro Act[6]. (16 U.S.C. § 1331). The statute recites Congressional resolution that wild horses "are living symbols of the historic and pioneer spirit of the West; that they contribute to the diversity of life forms within the Nation and enrich the lives of the American people…" (16 U.S.C. § 1331). However, the Act was so successful that by 1978 overpopulation was a serious issue and the statute was amended to empower the Department[7] to remove animals deemed in excess of the number the land could support without damage. *American Wild Horse Campaign v. Bernhardt*, 963 F.3d 1001, at 1004; 16 U.S.C. § 1333(a). This level is referred to as the "appropriate management level, the population level at which "the number of wild horses that can be sustained within a designated [area] which achieves and maintains a thriving natural ecological balance in keeping with the multiple-use management concept for the area," *Cloud Found. v. U.S. BLM*, 802 F.Supp. 2d 1192, at 1197 (D.Nev. 2011).

In the federal regulations that implement the statute, the agency is tasked with establishing and maintaining this level, and balancing competing uses of the land, in accordance with Federal

[4] Technically feral, but I will refer to them as wild, according to common and statutory usage, throughout.
[5] Wild burros also occupy these lands and are covered by the same statutory and administrative provisions. However, for the sake of brevity I will refer to horses throughout. As is often the case in the public eye, burros will be included by implication without mention.
[6] (PL 92-195, codified at 16 U.S.C. Chapter 30, also referred to as the Wild Horse Act, or WHA).
[7] Acting through its Bureau of Land Management.

land management statutes[8], and in consultation with other involved federal and state agencies, and "all other affected interests" (43 CFR 4700.0.1-0.2; 4700.0-6.) The regulation, while somewhat more detailed, is still by no means a step-by-step tasking and instead sets out broad responsibilities and parameters of authority. For example, the Bureau "shall prepare a herd management area plan," (4700.3-1). The details of what must be in the management plan and by what means it should be carried out, and when, are not specified. This leaves most decision making to the discretion of "authorized officials," meaning duly appointed employees of the Bureau of Land Management.

At the outset of the formal endowment of government authority—the first layer, so to speak—the original statute protecting the animals is the product of the federal democratic process. As such statutes often do, it uses broad strokes to authorize an executive-branch agency to use its administrative discretion in implementing that democratic product. The mechanics of exactly how Congress' intent to protect and manage the animals on public lands should be carried out is vague by necessity and design. The details are beyond the resources or knowledge of Congress to supply, while filling in such blank spaces is exactly what the administrative agencies are *for*, and they'd be rather powerless without it.

At the second layer, the implementing regulation is the result of that administrative agency's rulemaking process. It is mainly a product of agency discretion (within bounds authorized by Congress). Public notice and comment are required in agency rulemaking as a matter of course, but public comments have no binding force on the agency. This regulation, however, specifically requires that the Bureau consult "other affected interests," including interested commercial or private entities and persons, in carrying out its management plans. Most American citizens are probably not interested in (or aware of) the details of BLM's wild horse management program.

[8] Primarily the Federal Land Management Policy Act (FLPMA) and the National Environmental Policy Act (NEPA).

Those who *are* interested, however, are interested in passionate and particular ways. Even among those who construe themselves as advocating generally for the welfare of the animals tend to disagree deeply about what that welfare looks like. The agency officials exercising discretionary government authority in even the mundane workings of WHBP are not engaged in an activity on which public opinion is at all settled, even over issues to which it seems there ought to be a concrete factual answer (such as whether a particular piece of rangeland can in fact support the number of animals living on it, or whether it is or is not humane to treat the animals with both control, or to offer overpopulated animals for adoption by private individuals). Nearly every course of action the BLM pursues is likely to result in harsh criticism, and many result in litigation.

For example, in *Cloud Foundation* a wild-horse advocacy organization sought to enjoin a planned BLM roundup, in a dispute about how the regulatory provisions requiring that land be managed "principally" for the benefit of wild horses, and that this management be conducted at the "minimal feasible" level, ought to be interpreted.[9] "Principally for the benefit of," and "minimal feasible" level of management are highly subjective, vague terms amenable to wildly differing interpretations based on one's prioritization of competing land use, environmental, wildlife management, and other interests—including the rather sentimental concern, espoused by Congress itself, for maintaining free-roaming horses as "living symbols" of American history and spirit.[10] As

[9] The court, relying on the accepted legal principle that an agency's discretion exercise of discretion in interpreting its own regulations is subject to judicial intervention only where the agency has acted "arbitrarily and capriciously," found the BLM's actions to be reasonable and ruled in its favor—declining, as courts generally do, to wade into the often arcane details of a particular agency's area of expertise, under the presumption that this is why we have such agencies in the first place and that they are in the best position to make such decisions. This deference to agency expertise is called *Chevron* deference, originating in the Supreme Court's decision of *Chevron U.S.A., Inc. V. Natural Resources Defense Council, Inc.*, 467 U.S. 837 (1984).

[10] Of course, it is not only BLM's proactive efforts to control or reduce the population of wild horses that lands the agency in court. *Wyoming v. U.S. Department of the Interior*, 2015 U.S. Dist. LEXIS 189 (D. Wyo., 2015) was a suit brought by the State of Wyoming in objection to "final agency *in*action" (emphasis added), because it felt that the BLM had not removed *enough* wild horses from Wyoming public lands. That case, broadly, involved an argument as to whether the BLM's possession of discretionary *authority* to deem horses "excess" and remove them from the land, under the WHA, conferred a *duty* on it to conduct such removals or whether it instead has the discretion to determine that other considerations weigh against removal of animals in excess of an AML in certain cases

the long-established practice of judicial deference to the exercise of agency discretion recognizes, disputes about how these terms ought to be interpreted, or which agency course of action *best* satisfies them, may not readily admit of *any* objective or popular solution and yet a determination must nevertheless be made by the agency. This means that in administering the program, even low-profile, local level agency employees are at times exercising government authority to decide controversial issues. This individual discretionary decision making may raise objections on democratic grounds since its roots in a truly democratic process lie several layers of vagueness back. It is nevertheless necessary if government is to actually exercise meaningful leadership—lead citizens by giving guidance, resolving disputes, or setting priorities, for example—that is relevant to the routine lives of its citizens, and at least minimally efficient and responsive to their needs or interests.

Section II: Beerbohm's Relational Model

Eric Beerbohm proposes the "commitment theory" of democratic leadership, which he describes as a relational model, as a response to what he calls a tendency of other theorists to view leadership as somehow in tension with democracy, or even as anti-democratic. Beerbohm's goal is to give an account of leadership that explains how its exercise is consistent with democratic ideals. This section will discuss why the commitment theory has some important limitations. Specifically, the commitment theory does not adequately address the exercise of discretionary authority within the vast administrative state, and it also fails to address cases where meaningful commitment sharing is out of reach because of irreconcilable interests/commitments among followers, yet genuine leadership is nonetheless required.

According to Beerbohm, leadership is an interlocking structure of shared intentions (both between leaders and followers, and among followers) through which shared commitments to

(essentially, whether the agency may reassess whether the set AML actually corresponds to the level necessary to maintain ecological balance or not—whether and when "excess" animals are actually excess).

actions, orderings of priorities, etc., are formed. (2016, 643, 639). The relationship involved here between leaders and citizens is distinct from the phenomenon of responsiveness to voter preferences by elected leaders. (2016, 641). Commitment theory style leadership both balances and supports responsiveness in a democracy. (2016, 640). When this sort of leadership activity is successful, it takes on the character of a relationship (as opposed to mere one-way influence or command) in which citizens are "genuine participants" in "shared political activity" (2016, 639). There are two important functions involved that illustrate what Beerbohm means by shared activity, and which must be satisfied to constitute an instance of true leadership on that model - the commitment setting condition (meaning a leader's creating shared commitment among followers, or sustaining one that already exists), and the commitment mobilizing condition (meaning a leader's coordinating follower action towards that commitment). (2016, 642).

Beerbohm doesn't give an example of the full commitment theory in action, but one might go like this: suppose Senator X, from State Y has determined[11] that it is important for the continued economic welfare of her constituents over the coming 30-50 years to expand the volume of mineral leasing authorized on federal lands in State Y. In the first stage of commitment theory style leadership, she'll need to fulfill the commitment setting condition. It's likely that in this case there will be some constituents who already hold a commitment to Senator X's intended course of action, which she'll need to sustain. However, unless they already make up a majority of her constituency, she'll also need to establish that same commitment in others who do not currently hold it by convincing them both that it is important to take measures to ensure continuing economic welfare in State Y, and that increasing the volume of mineral and gas leasing allowed on State Y's Federal lands will be an effective way to do this. Assume that Senator X successfully sets (or sustains) this

[11] I have used "determined" for the sake of simplicity, whether it is the case that she has come to that conclusion independently, or been persuaded by either more-expert advisors, whether within the legislative or executive branches, or interest groups (expert or otherwise).

commitment among her constituents through a social-media based awareness campaign informing State Y voters of current negative economic projections for their State, and contrasting projections showing significant economic growth if more natural resource leasing is allowed. Then, moving on to the commitment mobilizing condition, those constituents themselves might do things like continue awareness-raising and commitment setting efforts to convince other State Y voters and representatives—or Senators and Congressmembers of other States—to support Senator X's proposal, organize efforts to comment publicly on any additions or changes to relevant regulations that are proposed by the Bureau of Land Management, or plan demonstrations aimed at pressuring elected officials to support Senator X's proposal. These constituents would then, as a result of their having come to share Senator X's commitment to expanding State Y mineral leasing as a way to ensure its economic prosperity, be carrying out political activity of various kinds with the goal of influencing the government to act in a way that is consistent with that shared commitment. Senator X has, as a leader, been instrumental in causing these constituents to take action, but that action itself cannot be directly attributed to her alone. The constituents are not acting solely because Senator X said they ought to; they have adopted the underlying commitment in a way that makes the decision to act internally motivated. This, I believe, illustrates Beerbohm's notion of leadership as a two-way relation resulting in shared activity, which he describes as having "a first-person plural aim: that we will do something together." (2016, 642).

To the extent that we're considering instances of leadership involving elected officials acting in cases where it is possible to create these kinds of shred commitments and coordinated efforts, Beerbohm's theory is convincing as a description of many cases of leadership, particularly by elected leaders. However, it fails to address two substantial, and sometimes overlapping, categories of decision making by government leaders, without which it can't constitute a *complete* theory of leadership within democracy.

First there are limitations resulting from Beerbohm's tight focus on elected leaders. These limitations matter because as I discussed in section one, an enormous share of actual government decision making happens within the administrative state. Then, there are limitations having to do with cases where interests—and therefore commitments—among constituents are irreconcilable, whether we're looking at a venue for action by elected or appointed-administrative officials. Within these two categories, there are three major points of limitation to discuss: two having to do specifically with the administrative state, and barriers to commitment-sharing rooted in both its nature and purpose, and one having to do with more general types of commitment-sharing barriers. It's important that we be able justify decision making in the sorts of cases covered by these categories where true commitment-sharing isn't possible, and yet leader action is still needed. We also need to be able to explain the appropriate constraints and considerations for how this power ought to be exercised to be normatively acceptable within a democratic framework. A theory that does neither is incomplete.

The first major limitation the commitment theory faces is that it doesn't translate well to administrative state decision making because of epistemic barriers that prevent true sharing of commitments (and possibly meaningful coordinated action). In shorthand, these are the Nobody Knows class of cases. These sorts of epistemic barriers—areas in which true subject-matter expertise may be needed to fully understand the issue to be addressed or problem to be solved, let alone identify an acceptable course of action—are one of the main reasons that the administrative state was created in the first place. This is exactly what agency experts are *for*, and why we give them so much discretionary authority within the bounds of their particular area of expertise.

The second major limitation is one that may overlap with, and sometimes even result from the first. It can also arise for reasons having nothing to do either expertise, however, and is at the heart of another important reason why we have so many, and such robust, administrative agencies in

the first place: barriers of attention or interest. These are the Nobody Cares cases. By this I mean that many administrative agency decisions involve such small or obscure stakes (or both) that most people don't know, don't have time, and often don't care about how these decisions ought to be or are in fact made. In the case of obscure, mundane, and low stakes decisions it may be impossible to generate the interest or attention. This isn't pejorative. It's plain recognition of a rational efficiency interest that factors among reasons that representative democracies, and administrative agencies, are desirable political structures in the first place.

The third major limitation of the commitment theory has to do with cases in which commitment-sharing is out of reach, or unrealistic, due to irreconcilable interests among constituents or the general unpopularity of (all) of the available courses of action; the hard cases, the Nobody Can Agree class of cases. Beerbohm doesn't say much about whether the success of this commitment-setting condition depends on any kind of critical mass by either proportion or absolute numbers. That is, whether the condition should be considered satisfied if even one constituent is brought to either sustain or share Senator X's commitment. However, since the second condition, the commitment-mobilizing condition, calls for coordinated political action it seems reasonable to conclude that there needs to be at least a large enough group in agreement to succeed to some degree in that action. Neither interpretive choice is particularly satisfactory, however. Let's return to the example case of State Y to understand why the issue of irreconcilable commitments may be an issue for the commitment theory as a comprehensive model of democratic leadership even outside of the administrative sphere. We can either conclude that the commitment theory requires that the commitment be shared by a majority of constituents (and not both shared and opposed by equal numbers) to count as a true instance of democratic leadership, or that it requires only that *some* number of constituents share that commitment such that Senator X does not act alone—no matter how small a minority that number may be. Either of these cases seems reasonably likely in a decision

implicating both the broad plethora of economic interests that depend on and benefit from the availability of federal mineral leases, and the environmental and other land use concerns related to the same. The first case reduces to simple majoritarianism and seems insufficiently distinct from responsiveness. The second case seems anti-democratic in exactly the way Beerbohm seeks to avoid, if our metric is to confer normative approval on leadership actions just in case they involve shared commitment and coordinated action between a leader and any number of constituents.

The commitment theory accounts for only one sort of leader-action (setting commitments and mobilizing group action), without giving adequate attention to cases calling for decisive leader action. He acknowledges that at times a leader will be required to decisively choose between two (or more) equally-rational courses of action by comparing the leader-as-tiebreaker's role to Raz's account of decision making in such rationally underdetermined cases. He notes that these cases where the reasons have run out may help to define our identity as a people in the same way that Raz says they define our identities as persons—but he seems to conclude that a leader will handle such cases by convincing followers to share her non-rational commitment to whichever course she has chosen and then simply moves on. (2016, 643). This is far too optimistic. Some commitments simply are not going to be widely shared, and some will be opposed just as enthusiastically, and in similar numbers, as the one a leader has chosen. To conclude that there is something going on in these cases of unpopular or opposed commitments that makes leader action specially democratic *by virtue* of their being shared by *someone* seems unjustified, but such cases are unavoidable for government decision makers at all levels. These cases are important, and they can be accounted for without concluding that they are either 1) anti-democratic in the way Beerbohm and others seek to avoid by highlighting commitments, responsiveness, or the like, or 2) something other than "true instances of leadership." In fact, they are precisely the kinds of decisions which we most need leaders for.

These sorts of irreconcilable interests are certainly in play in the example from section one regarding BLM's wild horse management responsibilities, just as we could expect them to be in the example of State Y and its mineral and gas leasing arrangements. There is a degree to which the interests involved are fundamentally incompatible, and it isn't rational to expect that the parties involved could be expected to share a commitment regarding how the government ought to act. People making their living raising cattle that graze on public land cannot be expected, generally, to share a commitment towards prioritizing forage and water on those lands for the use of wild horses. People invested in not only the welfare, but also the continued free-roaming nature, of those horses (as many wild horse advocates are) cannot rationally be expected to sincerely share a commitment to limit the population or range of horses for the sake of accommodating grazing leases. Nobody can agree. To expect that a leader must cause such a fundamental change in commitments in either group in order to truly be exercising democratic leadership demands too much of theory and perhaps too little of practice.

Section III: Sabl's Theory of Democratic Constancy as a Guide for Leader Behavior

Andrew Sabl approaches the question of how leaders—elected and otherwise—ought to act within a democracy by seeking normative principles that can guide their activities according to which particular office they occupy. (2002). Much like how Beerbohm tells us important things about the nature of at least one aspect of democratic leadership—that it often has the character of a relationship between leaders and followers, and among followers—Sabl tells us some important and useful things about characteristics or decision-making habits that we may desire in those leaders. However, his theory too is incomplete in that it overlooks the importance of the discretionary power exercised in the administrative branch, and cannot adequately guide leaders in epistemically intensive, obscure, "irreconcilable interest" or otherwise underdetermined cases.

For Sabl, a political "office" may be an official role in a legislative, judicial, or administrative branch of government or as an executive, but it may also be a role as an activist, community organizer, or possibly some other unofficial capacity. (2002, 11-12). After concluding that neither particularist accounts (which hold that it is acceptable for officeholders to act with particular preference for certain groups, causes, or principles) nor universalist accounts (which would reject all such exercises of preference) offer a satisfactory guide for political conduct, Sabl concludes that the answer lies in compromise between the two. (2002, 55). He finds this compromise by tempering the desire for the application of some sort of consistent moral standard with the recognition that well-functioning and flourishing democracy is a matter of institutional health and robustness. He proposes that we ought to divide moral responsibility according to a person's office. (2002, 55-56). This means that the moral standards by which we judge officeholders will not be entirely particularist, but they *will* differ according to which office is held in that the underlying moral requirement is, in all cases, whether the officeholder in question is acting in a way that causes her to serve

Whatever behavior constitutes serving well in a particular officeholder's specific role, then, is what constitutes moral behavior for that officeholder's that role. This in turn means that moral behavior for political officeholders is whatever behavior contributes to a well- or better-functioning set of democratic institutions. The degree to which a particular officeholder is more or less concerned with specific democratic ideals (such as liberty or equality), or institutional practices (such as judicial norms or the principles of fairness in the electoral process underlying campaign finance regulations) will differ according to which office that person holds. (2002, 56). Sabl proposes that the particular attachments people occupying various offices may have to particular constituencies (if they are elected representatives), causes (if they are activists), or institutional structures (for those holding official office - likely especially those in appointed offices in the judicial or administrative

branches) are actually not threats to good, fair governance at all but instead motivating factors that lead these officeholders to perform well and thereby safeguard healthy political function. (2002, 55).

When it comes to the mechanics of how he envisions this safeguarding to occur, Sabl reminds us that democracy is ultimately reliant on the proper function of institutions. In setting forth standards for how officeholders ought to act in order to best safeguard these institutions he focuses on the question of what character we would wish for them to demonstrate in carrying out their duties, drawing inspiration from Cicero to propose that what is required is a virtue of constancy. (2002, 56). By constancy he means a steady attachment to values, principles, and goals in the long rather than short term. (2002, 56). This sort of steadiness is valuable to the democratic system, he says, because such a system is characterized by the denial of many short-term desires (perhaps my wish that senator X vote for issue A rather than issue B) in favor of long-term goods (perhaps my wish that my government function in such a way that my interests as well as those of my differently-situated neighbors are considered in some way that is at least minimally equitable.) (2002, 57). This means that good democratic officeholders will both "refrain from acting in the face of temptation" (whether born of self-interest, or the wish to fulfill some short-term desire of a group or cause to which one is particularly attached, or some other temptation), "but also the positive ability to continue to act as one has resolved in the face of dissuasion, opposition, or even hatred." (2002, 57). An official who possesses Sabl's character for democratic constancy will be guided to better decision making, especially, in those cases where conflicting passions are high and the temptation of a quick—or short-term popular—fix in play.

Sabl gives detailed examples of the principle as applied to three types of office which he finds to be of special interest and importance - senators, activists, and organizers. He notes in passing that administrative offices may present ethical problems for decision makers, whether executives or not, but concludes that issues surrounding these offices largely fall under the umbrella

of the well-discussed "dirty hands problem," (politicians doing the wrong thing for the right reasons), and that these hard cases tend to sway our perspective and distract our interest from the day-to-day "less-dramatic" decision making that happens in other offices. (2002, 2-3). In doing so, he makes a mistake similar to Beerbohm's by exhibiting the same general failure to assign adequate significance to the impact of administrative agency decision making and its importance for a full understanding of how government authority is, and ought to be, exercised in a democracy.

Previously, I explained how my three classes of administrative state problem cases for government decision making are not adequately addressed by Beerbohm's commitment theory of leadership, in large part due to this oversight. These were Nobody Knows cases featuring specialized epistemic requirements for successful decision making, Nobody Cares cases of such obscurity, complexity, or low stakes as to present little or no interest to the public, or even most government officials, and Nobody Can Agree cases where conflicting interests are irreconcilable or all available courses of action unpopular (or both). Sabl's focus on conflicting interests, and how an officeholder ought to deal with both her own and others' attachments to them, does take better account of the challenges presented by the third category of Nobody Can Agree cases in that section. However, in all three of these categories the theory of democratic constancy still offers no assistance for resolving *what* an officeholder exercising administrative discretion ought to decide, beyond perhaps the vague endorsement of her having a particular attachment to the integrity of institutional procedures. The theory offers only very general assistance as to *how* she ought to decide: presumably by honoring those institutional procedures, and by exhibiting democratic constancy in prioritizing long-term institutional health over a short-term desire to make an expedient, popular, or self-serving decision.

Consider an example of an epistemic-barrier Nobody Knows case. Suppose that Senator X of State Y's office has received numerous complaints from State Y businesses that federal regulations for the frequency with which heavy mining equipment must receive safety certification is

unreasonable with regards to a particular drive-shaft component used in boring machinery because it requires that work stop for the duration of the inspection and results in significant loss of profit. The geography of State Y is such that this particular type of equipment is necessary to successful mining operations in most of its territory, whereas it is unnecessary and therefore very uncommon in other states. Senator X's office then sends a pointed inquiry to Agency Z, wanting to know why the standard has been set in a way that harms economic interests in her state; Agency Z is then required to testify before Congress to defend its standards, and political pressure mounts for their revision because Senator X's support is instrumental to the success of an unrelated bill that a majority of her colleagues hope to pass this term, and which the President supports. The Whitehouse has requested that Agency Z review the inspection standards to determine whether the current annual frequency is necessary to ensure acceptable standards of safety, or whether a less-frequent schedule would suffice. Official Q at the agency, an expert on the safety standards in question, must then decide whether a revision to the regulation is appropriate. He discovers during the course of his research that studies pertaining to the frequency of inspections for that equipment are split 50/50 as to whether the current annual standard should be required, or a biannual frequency is adequate instead. Meanwhile, several interest groups have taken notice and begun organizing movements in support of maintaining the current safety standards for the protection of workers who could be injured in the case of a malfunction. Official Q himself happens to be the uncle of an equipment operator who was injured in an accident involving a faulty drive-shaft 10 years prior. There is no clear epistemic basis according to which he ought to decide. The only available factor which might tip the scales are his particular attachment to his injured family member, or the consideration of whether it is more undesirable in the short-term for an agency to displease a Senator or a collection of presumably-powerful interest groups. Democratic constancy will help Official Q avoid the temptation of a quick-fix decision in favor of either side simply for the

sake of getting them off his back. It may even help him, through his particular attachment to the proper functioning of the administrative agency as an a-political and publicly interests democratic institution, to avoid a decision based on his own personal attachment. But it will go no further, and the same would be true if the decision were instead one which had garnered no public or Congressional interest at all, or one in which the opposing interests (in generating steady mining revenue, versus protecting the operators of heavy mining equipment) were fundamentally irreconcilable.

Like Beerbohm, Sabl offers little help for the discretion-holding leader who must take decisive action, and be able to justify said action, in the absence of convincing reasons on any side (the cases where "reasons have run out"). If reasons have run out, no amount of steady dedication to institutional function or democratic ideals will provide them where they do not exist. Likewise, though particular personal attachments *might provide such otherwise-absent reasons*, a decision made on such a basis would certainly fail Sabl's test for moral, desirable officeholder conduct: we would not conclude that an administrative official's acting in such a way could possibly enhance, or even fail to undermine, the health of the democratic political structure overall. It wouldn't be an act that could be considered "serving well" (Sabl, 2002, 55) in her office. Our leader is again left with a vacuum of reasons, seeking to make a decision that is still principled, moral, and *reasonable*.

Section IV: Heath's Ethos of Public Administration as a Policy-Setting Guide

Unlike Beerbohm and Sable, Joseph Heath recognizes the importance of administrative-state decision making. Heath acknowledges that administrative authority can be a vital feature of how and why modern democratic states function successfully. He even highlights the fact that citizens often view such administrative bodies as *more* reliable or effective than elected officials. (Heath, 2020, 13). Heath also acknowledges the aspects of such authority that may cause administrative power to appear at odds with democratic ideals, given that it involves the discretionary exercise of significant

power by unelected officials, often bounded only by vague or permissive statutory mandates from the elected branches and after-the-fact judicial oversight that tends to treat administrative decisions with deference. Heath concludes, though, that we can identify normatively appropriate boundaries for these exercises of authority within a system of democratic power in two ways. First, we should understand their role and position within the institutional checks and balances of the modern liberal state (perhaps strengthening such institutional constraints where necessary), and second we can identify the underlying principles of normatively appropriate governance that are already relied upon by these officials as an informal such set of boundaries (and again strengthen such principles, as well as adherence to them, where necessary). (Heath, 2020, 45). Heath's theory, like Beerbohm's and Sabl's, provides helpful insight for justifying the existence of administrative discretion as a general matter. It even speaks to a particularly more robust version of administrative authority by considering how far-reaching agency impact on the sort of policy-setting decisions that are often left to elected officials in the U.S. system. However, it also fundamentally relies on our acceptance that active promotion of public interest within the framework of the liberal welfare state is an acceptable and desirable function of government (a position that at least some are likely to reject) for its justification of administrative authority. Furthermore, it still falls short as a complete theory of individual-leader decision making, or even administrative authority, because it fails to concretely address the same three classes of cases which present problems for the other theories: the Nobody Knows, Nobody Cares, and Nobody Can Agree cases.

Heath is primarily interested in constraints on how administrative agencies exercise power that can influence policy-setting at the highest levels (as opposed to lower-level or lower-stakes decisions more directly involving individual members of the public). (Heath, 2020, 19). His understanding of the administrative state is as a moderating and counter-majoritarian force within government (balanced particularly against the elected legislative branches), which is ironically not

unlike the role that Sabl envisioned democratic constancy might guide Senators to assume. (Heath, 2020, 11-12). He refers to the set of experienced career officials powerful enough to have this type of influence as the permanent civil service.[12] Heath argues that these high-ranking individuals act in accordance with an understanding of the public interest that is independent of, and may even conflict with, that accepted by elected officials. In doing so, he says, they perform a function that is not only consistent with traditional democracy, but important for a well-functioning modern liberal democracy. (Heath, 2020, 345). This concept of public interest is not simply majoritarian. It is also not necessarily consistent with the sort of legislative responsiveness (to the desires of constituents) that Beerbohm discusses.

Heath's project, then, is to explain how this use of authority—and especially its influence in shaping policy--can still be consistent with traditional democracy and avoid reduction to mere domination or "bureaucratic tyranny". (Heath, 2020, 19). The two main ways a system of government can avoid these pitfalls are to ensure robust explicit mechanisms of institutional transparency and accountability as part of a "rule-of-law culture" (Heath, 2020, 290) are in place, and to promote a culture of adherence to a set of "public service values" as implicit, informal rules to guide official action in areas of administrative discretion where laws or regulations do not otherwise provide guidance (Heath, 2020, 346). Heath advocates for an "ethos of public administration" that provides a broad guide for the way in which public officials ought to make their determinations, similar in some ways to Sabl's account of desirable character traits for public officials. This guidance is intended to ensure that their authority remains normatively democratic, by prompting them to act

[12] Heath is primarily focused on parliamentary systems that do not share the U.S. practice of reserving certain influential or high-level positions for appointment directly by an elected official, thus increasing the independence of the permanent civil service from any current crop of incumbent elected officials—and also heightening the need to explain how its autonomous exercise of authority remains nevertheless democratic. Many features, however, remain in common between these systems and the U.S. system, and most of the decisions this paper is particularly concerned about are, in that system, made by what are called "career," as opposed to politically appointed, officials.

with concern for the principles of "efficiency, equality, and liberty which form the normative core of modern liberal theories of justice." (Heath, 2020, 45).

Heath's discussion is most concerned with the administrative state as it manifests in the liberal welfare state--hence the focus on possibly counter-majoritarian notions of the public interest. He focuses on this invocation of an administrative idea of public interest in service of various exercises of authority by a modern welfare state in setting policies whereby non-elected officials exert government authority over us for the sake of "our own good," [13] and whether or when they can be adequately justified. (Heath, 2020, 344). The decisions of administrative officials acting to set policies that they have deemed to be in the public interest independently of elected official authority are normatively appropriate (and more likely to be justified/less likely to be coercive in unjustified and anti-democratic ways) when those decisions are just. Likewise, the public interest that officials ought to promote in exerting this kind of influence lies in an outcome that is consistent with justice understood as efficiency, equality, and liberty. This respect for individual liberty interests will in theory guide Heath's officials to set policies that are paternalistically coercive (and therefore in tension with democratic liberty ideals) only where they can overcome the high justificatory burden required by this understanding of justice and are therefore consistent with the modern liberal democratic welfare state.

[13] Heath provides a detailed discussion of various arguments against paternalism, generally concluding not that paternalistic policies ought to be rejected merely on the grounds that they are paternalistic, but instead that this characteristic increases their justificatory burden because they are more likely to be coercive in nature. (Heath, 2020). This willingness to weigh paternalistic authority against potential justifications, and thus possibly defend at least some degree of paternalism, is a feature of the theory that seems consistent with its willingness to embrace an equation of modern administrative authority that is in support of the public interest (and therefore democratically legitimate) with the liberal welfare state. I do not think, however, that it is necessary for the acceptance of the general claim that public servants can exercise significant government authority independently of elected officials in a way that is democratically legitimate, and that in doing so they ought to be-and are-constrained by some general ethical principles having to do with ideas about what is in the public interest. I have therefore declined to address the question of paternalism and Heath's detailed account of the rise of the welfare state in any greater depth here.

The questions of how individual decision makers ought to make individual decisions, or of how we ought to understand leader discretion within democracy as a general matter, are not questions of primary interest in Heath's work. Because of this, it is not entirely surprising that he doesn't say much about *exactly how* this "ethos" might guide individual decision makers to make (normatively, democratically) appropriate individual decisions. However, these problems remain both present and important. The public service ethos alone isn't enough to guide us in the Nobody Knows, Nobody Cares, or Nobody Can Agree cases: the same cases in which, for Beerbohm, we would struggle to explain how a leader might successfully set a commitment to be shared with and among followers, or for Sabl how a character for democratic constancy alone will not supply reasons that do not exist. In Nobody Knows cases featuring epistemic obstacles, even experts may be unable to clearly identify one or the other decision that is guaranteed to best serve any particular idea of the public interest because of insufficient evidence or difficulty in making reliable predictions, or even deep disagreements among experts about how what information *is* available ought to be interpreted. In Nobody Cares cases of obscurity, complexity, or low stakes it will often be the case that the public *isn't interested* specifically because there is so little impact on "the public interest." In Nobody Can Agree cases of irreconcilable interests or equally unpopular options, we may be unable to say which set of interests should count as public, or which unpopular option is unpopular for the right reasons, and which is unpopular but nevertheless for the public's "own good." Furthermore, we may be unwilling to accept that this sort of independent identification of what is in the public's interest—lacking legislative or popular guidance on the matter—is always a desirable approach.

Meanwhile, Official Q must still determine whether to send a report to Congress that concludes that the current one-year inspection frequency for mining equipment drive-shaft safety is necessary or one that concludes the frequency can safely be extended to a two-year interval, and the Wild Horse and Burro Program must continue to set appropriate herd management levels, and

propose gather and population control measures in the face of public opposition to all available courses of action. Heath's public service values may help our officials rule *out* certain patently inefficient, or self-serving courses of action but they will not help to rule *in* one among a number of equally reasonable, equally appealing, or equally unpopular options. They provide another formulation of established principles that discretion ought not be exercised in a way that is simply arbitrary, unreasonable, or self-interested and general features that administrative discretion should demonstrate to best avoid accusations that it is anti-democratic. As useful for some purposes as this may be, it doesn't help to determine how, in underdetermined cases, arbitrariness and its fellows are best avoided.

V: Making Oneself Accountable

In the preceding sections, I have set out an account of a relationship between vagueness and discretion in the law, and its role in the structure of the modern administrative state, that highlights the need for a theory of how individual leaders both do and ought to exercise government authority in what I have referred to as a variety of classes of underdetermined cases. I've explained how Beerbohm's model of leadership in democracy is descriptively incomplete because it doesn't capture these exercises of individual discretion, and normatively incomplete because it therefore doesn't speak to how they might be made acceptably democratic. I've also discussed Sabl's account of ethical democratic leadership in a variety of capacities (which he refers to as offices), and how it too neglects this important area of government decision making. Finally, I've addressed Heath's effort to speak exclusively to the question of how and why administrative power can be properly democratic, which recognizes the importance of this particular type of governmental authority but again falls short of speaking in detail about how an individual decision maker ought to make use of her discretion in the hard cases.

There are then two important questions to answer. First: if Beerbohm's commitment theory

doesn't capture all instances of normatively acceptable leadership in democracy, because it cannot capture the activity of the administrative state, what *does* describe this sort of leadership activity? Second: what are the characteristics of normatively and politically acceptable decision making for these types of leaders that Sabl and Heath have not provided?

This is not to say that these three proposals do not have anything to offer on these questions, even if they have not themselves answered them. From Beerbohm we gain the insight that what we would refer to as genuine leadership, as opposed to domination or pandering, has a relational character. From Sabl we gain the insight that democratic governance is importantly about the functioning of institutions more than the discrete acts of individuals, but that those institutions function at their best when the individuals within them carry out their particular offices well, in whatever way may constitute "well" for the office in question. Sabl's focus on best possible outcomes and long-term institutional health is also an important guide for leaders facing hard cases. From Heath we can take a recognition that un-elected officials who exercise discretionary authority within democracy must 1) be in some way properly accountable to other democratic institutions (and the public) for their use of this power, and 2) be not only subject to, but adopt of their own volition some sort of ethical code constraining their discretion in cases where statutes or regulations do not.

To answer my two remaining questions, I propose a theory of leadership for discretionary decision making that is characterized by the act of *making oneself accountable*. The act of making oneself accountable is distinct from the state of *being accountable*, both practically (because others, possibly including the public, will know a decision has been made, and who is accountable for having made it) and institutionally (because one might be *held accountable* to supervisors, disciplinary boards, elected officials, or the judicial system). The distinction between the active and passive here is substantive. A leader who *is* but does not *make* herself accountable may act or refrain from acting in

a particular way for fear of being held passively accountable to others, or simply without regard to the possibility or necessity of external judgment. This sort of action may even lead in many cases to desirable results, but it may fall short at the very least in cases where there is no externally mandated or prohibited course of action, or in which an actor believes (rightly or wrongly) that her acts will escape scrutiny. It may also fall short in cases where a later dispute, or the need for subsequent similar or related decisions requires an accounting of the original act which is now not available. If a leader actively *makes* herself accountable these cases disappear, both since the scrutiny to be avoided would in the first instance be her own, and because in the second instance she will take steps to document and where necessary publicize any acts that would otherwise remain unobserved. The act involves adherence to three related principles.[14] I will explicate these three principles in turn, illustrating them with reference to the case of Official Q, who was charged in the previous section with evaluating safety inspection guidelines for potentially dangerous mining equipment.

The first principle runs as follows:

> 1. *Deliberate Assumption of Risk*: an official exercising discretionary authority ought to do so with a deliberate awareness that the decision could be the wrong one, or at the least could have bad results for some people.

To meaningfully assume the risk that accompanies their authority, officials should act with conscious awareness of the possibility that they may make the wrong decision, including at least some reflection as to who might be affected. Leaders should not merely identify which course seems to be correct or desirable and hope for the best, they should consciously recognize the possibility that if they choose poorly, the worst may occur because of their decision. In section three, I presented the example of Official Q, who has been tasked by the Executive with determining whether potentially hazardous heavy machinery must be subject to disruptive and costly inspection annually, or less

[14] These principles are intended to apply to the bulk of individual leader discretionary decision making. However, it is important to note that emergency situations could alter our assessment of whether they have been adequately applied. That discussion, however, is beyond the scope of this paper.

frequently (assuming, for the sake of discussion, that a response of "inconclusive" is not an option, since he has been specifically required to provide a recommendation). The facts available are split as to whether the current annual schedule is necessary to prevent excessive safety risks, with industry and public opinion also failing to break the empirical tie. Official Q risks, on the one hand, a recommendation that could cause the injury or deaths of machine operators (among other bad results of any equipment failure) if he decides to recommend less frequent inspections. However, it is not possible to be certain that such failures would in fact result, and if he recommends that the annual requirement be left in place he risks unnecessarily prolonging a heavy economic burden on the industry affected (and as a secondary effect, turning public and industry opinion against the administration, at least in certain segments). Official Q is obligated to identify all of these risks, and accept them, before making a decision that is informed by them.

The second principle of my theory of leaderships is as follows:

> 2. *Dedication to the Public and the Institution*: A leader must exercise discretion in a manner that is intended to promote the best-identifiable interest of the population to be served, and the continued health and respectability of the institution on whose behalf the leader acts.

This principle signifies that decisions ought not to be self-interested (or biased in favor of a particular group without identified and documented reason), reckless (lacking a *deliberate* assumption of risk, since a leader should act knowing that the best-identified interest may not be in fact the best), or arbitrary. Self-serving or unjustifiably biased decisions are ultimately bad for both citizens (for obvious reasons), and institutions (because they erode public trust and ultimately undermine government's ability to function). This principle prioritizing transparency and service of the public trust is neither novel nor unique, because it is a baseline requirement for any reasonable effort to describe a defensible exercise of democratic political authority. The important feature for my purposes is the invocation of transparency in service of my first principle, assumption of risk. To fully assume the risk associated with power, our decision-maker must act in a way that makes her

intentions and process transparent so that she can be fully accountable. Any attempt to avoid transparency would be an attempt to avoid the full risk associated with her decisions. Adherence to this second principle is, in part, a component of how a leader can comply with the first. This means keeping detailed records in a public system of how a leader has arrived at his decision. In addition to a recounting of the available facts that were weighed, those details should include how the relevant risks accepted pursuant to principle #1 were identified, and how they were weighed, so that a leader's reasoning can be scrutinized in the future if necessary. For Official Q, this would mean a recounting of the available facts, including an explanation of why they do not resolve the question, along with a listing of the relevant public interests (industry, versus local economic/employment and environmental concerns, versus worker safety), along with Official Q's conclusions about the relevant risks and benefits of either potential recommendation. It must also explain why he has chosen to make whichever recommendation that he has. To do a satisfactory job of giving this explanation Official Q will need to turn to the third principle.

The third principle runs as follows:

> 3. *Identification and Ordering of Priorities*: A leader must identify the various interests affected by the decision to be made, and document which are to be given priority over others and why.

This principle says that a leader should either identify the relevant principle that prompts a particular ordering of priorities, or supply one—and a justification for its selection—when one does not already exist. This should be supported through careful record keeping and transparency measures. This principle supports the first, because in order for a leader to fully assume risk for a decision we must be able to know not only what was done, but why. Perhaps more importantly, however, in cases where there is no best-identifiable interest to be served (whether that is because none exists, or it simply can't be identified), or it is not clear which course of action will best serve that interest, it avoids the arbitrary by requiring a leader to look outside of the immediate decision instance and supply reasons where there seemed to be none.

In Official Q's case, this will require that he identify why certain risks or benefits associated with one of the recommendations he could make (annual or less frequent inspections, given that the facts do not conclusively establish whether less frequent inspections will be just as affective at preventing equipment failure, or will be likely to result in more frequent failure) should be weighed more heavily. Since he has been tasked with providing a recommendation instead of merely a report, this requires that he make determinations about whether, for example, worker safety is to be prioritized over regional economic prosperity. In many cases this will mean Official Q falls back on existing policy to find some evidence of which public priorities exist, and how they should be ordered, but he will have to actively seek the information out and it may not always exist. Where it does not exist, he will have to independently establish such an ordering, and justify it on the basis of some other existing regulatory or legal principle.

Suppose, for example, that Official Q needs to place the risk of environmental damage from an equipment failure into context among concerns about worker injury and the cost of stopping work for inspections. He[15] might turn to overarching Congressional expressions of intent for environmental policy that can be found in the National Environmental Policy Act (NEPA) for guidance in explaining how and why those particular risks are given the priority that they are, but since the guidance is inconclusive on the facts of our example case, he will still be responsible for stating where in the ordering that particular priority lies. This exercise requires Official Q to place his use of discretionary power solidly within the context of available democratic expressions of preference, even if he must sometimes provide those preferences himself. Recall that in the initial presentation of Official Q's task we learned he himself had a close family member who had been injured by the failure of the type of machinery in question. There is a wildly important difference

[15] Likely with the assistance of an army of career federal Attorney Advisors in the agency's general counsel's or solicitors' office.

between his saying that "existing expressions of intent from regulators and Congress indicate that for reasons x, y, and z worker safety ought to be prioritized over the economic impact of inspections," "worker safety ought to be prioritized over economic impact" (without provision of any reasons), and "worker safety ought to be prioritized because no one else should be injured the way my nephew was." The last is clearly unacceptable. The second fails to make itself intelligible in the way the three principles require, so that we do not know if the decision has been made for acceptable reasons or not. The first respects the principles, making Official Q accountable in the appropriate way. These sorts of priority identification and weighting explanations, taken together, will begin to shape a previously unavailable framework of reasons by which Official Q can now explain a decision that might otherwise have appeared to be arbitrary.

Avoiding arbitrary decision making is a widely accepted baseline requirement for anything approaching good governance. It is a central principle of U.S. administrative law, incorporated into the Administrative Procedure Act at 5 U.S.C. § 706(2)(A), and underpins various theories of public justification in political philosophy, requiring that the use of government authority must be publicly justified on the basis of publicly available, or even publicly endorsed, reasons, though this presents deep problems for the administrative state which often acts for reasons entirely inaccessible to the general public (Rawls 2005, Christiano 2004, Kogelmann and Stitch 2021).[16] This does not mean, however, that administrative decision-makers should be let off the hook by way of impossibility, or that they are in fact necessarily engaging in an activity at cross-purposes with most notions of properly justified democratic power. Providing such justifications to the greatest and most publicly

[16] Rawls (2005) requires reasons that the public can endorse to justify the use of (necessarily) coercive political power. Christiano (2004), discusses a requirement of "weak publicity" as a principle of (social) justice. Kogelmann and Stitch (2021) argue that a Rawlsian requirement for public endorsement of reasons is incompatible with administrative decision making because the administrative state exists specifically to make decisions for which the kind of reasons demanded by the theory do not exist. (1-2, 12, 26).

intelligible extent practicable is in fact one half of the particular duty held by these leaders (identifying the non-public reasons on which they rest being the other).

VI: What to Do When Reasons Run Out

I have argued that discretionary decision making by individual leaders is a necessary feature of modern democratic government for reasons having to do with the structure of such laws and institutions and the vagueness inescapably incorporated into them. The administrative state is the venue for most such decision making, and therefore presents important and unique problems for theories of democratic authority and leadership which should not be ignored. I've discussed three such theories, and shown where they struggle to give a useful framework for how power is or ought to be exercised in three sorts of hard cases frequently facing administrative leaders: 1) those featuring specialized epistemic requirements, 2) those of great obscurity, or complexity, or in the disaggregate such low stakes, that they present no public interest, and 3) those where the public interest *is* captured, but those interests validly compete in irreconcilable ways or where all potential courses of action are publicly unpopular.

The principle of making oneself accountable offers leaders a coherent method for handling such cases in a way that takes obligations to democratic ideals seriously while also respecting the need for decisive individual action in the absence of clarity about public mandates, guidance from elected officials, or "the" unified public interest (and without resorting to apologetics for paternalism). As in the example of Official Q's decision, a leader facing any one of these three sorts of cases will engage in a process of explaining the factors that guided her decision, whether that means referencing whatever specialized knowledge weighed in favor of the chosen course and documenting where and how it was used (even if it is not possible to explain the fine details), or accounting for how those various irreconcilable interests were weighed against one another and why one was given priority over another (for some reason more substantial than a coin flip).

Making oneself accountable for the use of government authority, then, requires an awareness that one acts in the context of a relationship with those whom authority is exerted over (recalling Beerbohm), that one ought to approach this relationship with certain types of attitudes and not others (reminiscent of Sabl's characteristic for constancy) and in accordance with an articulable ethical code (reminiscent of Heath's public service ethos, or the United States' Code of Federal Ethics for employees), and that one's role as a leader fundamentally requires taking the initiative to identify or even create priorities and impose order on the chaotic and indistinct. Whether an official fully engages in the act of *making herself accountable*—rather than merely *being* passively accountable—tells us important things about her attitude towards her own authority that allow us to assess whether she has acted in a way that is democratic or despotic even in cases where we can't rationally claim that one course of action objectively ought to have been chosen over another. This active effort distinguishes good-faith from fiat, making what could be domination and bureaucratic tyranny into an indispensable means whereby government is able to effectively function "for the people" in areas where popular opinion is not an appropriate guide.

Works Cited

Asgeirsson, H. (2015). On the Instrumental Value of Vagueness in the Law. *Ethics, 125*(2), 425-448

Beerbohm, E. (2015). Is Democratic Leadership Possible? *American Political Science Review, 109*(4), 639-652

Boyd, D. and C. S. Boyd. (2019). *Ecological Effects of Free-Roaming Horses in Borth American Rangelands, BioScience, 69*(7), 558-565

Christiano, T. (2004). The Authority of Democracy. *The Journal of Political Philosophy, 12*(3), 266-290

Congressional Research Service (2021). *Federal Workforce Statistics Sources: OPM and OMB.* Retrieved from https://sgp.fas.org/crs/misc/R43590.pdf

D'Amato, A. (2011). On the Connection Between Law and Justice. *Faculty Working Papers.* Paper 2. Retrieved from http://scholarlycommons.law.northwestern.edu/facultyworkingpapers/2

Hart, H. L. A. (1958). Positivism and the Separation of Law and Morals. *Harvard Law Review, 71*(4), 593-629

Heath, J. (2020). *The Machinery of Government: Public Administration and the Liberal State.* New York, NY: Oxford University Press

Kogelmann, B. and S. G. W. Stich. (2021). When Public Reason Falls Silent: Liberal Democratic Justifications versus the Administrative State. In D. Sobel, P. Vallentyne, and S. Wall (Eds.). *Oxford Studies in Political Philosophy Volume 7*. New York, NY: Oxford University Press

Plato (1999). *Statesman*, Christopher J. Rowe (Trans.). Indianapolis, IN: Hackett Publishing Company

Sabl, A. (2002). *Ruling Passions: Political Office and Democratic Ethics.* Princeton, NJ: Princeton University Press

Waldron, J. (1994). Vagueness in Law and Language: Some Philosophical Issues. *California Law Review, 82*, 509-540

Cases and Authorities

Am. Wild Horse Campaign v. Bernhardt, 963 F.3d 1001 (9th Cir., 2020)

Chevron U.S.A., Inc. V. Natural Resources Defense Council, Inc., 467 U.S. 837 (1984)

Cloud Found. V. United States BLM, 802 F. Supp. 2d 1192 (D. Nev., 2011)

The Wild Free-Roaming Horses and Burros Act of 1971, 16 USC Chapter 30, Public Law 92-195

Wild Free-Roaming Horse and Burro Management, Code of Federal Regulations, Title 43, Subtit. B, Ch. II, Subch. D, Group 4700

Wyoming v. United States DOI, 2015 U.S. Dist. LEXIS 189415, 2015 WL 12916334, (D. Wyo., 2015)

Paper 2: Seeking Standards for Leadership Reasoning in the Executive Branch by Analogy to Representation and Judicial Reasoning

Each branch of government[17] serves a distinct purpose, and each has a type of activity that is most characteristically associated with it, but government activity as a whole can be broken down into a series of categories that correspond not with the branch where they take place but instead with the relationship an official carrying them out has in that instance with the law. Officials may make, interpret, or apply law. At times, however, they may also do something else, characterized by their acting in accordance with a grant of discretion under the law to fill gaps left behind in what had been made or interpreted, which is therefore available to be applied. None of these types of activity are the exclusive purview of one branch, although which of these types of activity an official engages in is influenced by what branch they belong to, and each branch is characterized by one type in particular. The standards for how we should evaluate decision making throughout these activities will differ according to a number of factors, including what larger institutional purpose the decision was made to serve. In the case of the executive branch, its most important purpose is not to mechanically apply the law the legislature has made or the interpretations the judiciary has provided, but to employ discretion to fill those remaining gaps. The standards by which we evaluate an official's decision making differ according to the institutional purpose they are expected, by role, to serve and the context in which their decisions are made. This means that while we may be able to draw parallels from standards for decision making in the legislative or judicial branches, neither of these contexts can adequately tell decision-makers in the executive branch how to gap fill even where the process of discretionary reasoning involved is the same. However, understanding that

[17] I specifically discuss the tripartite United States system in this paper, but the basic rationale would apply to any system with a distinct body of executive or administrative officials—bureaucrats and military officials alike—empowered with discretionary authority to make binding government decisions outside of the legislative or judicial processes.

framework as whole, and the place of discretion within it, is critical to understand why the types of decisions they are empowered to make render gap-fillers necessary, and what we should expect from them.

The activity discussed in this paper concerns the work of lower level administrative and military personnel, not the decisions made by the President or cabinet-level officials. We should not discount the importance of so-called "petty bureaucrats," since those insignificant low-level workers can be important government leaders when they play their gap-filling, and this role is crucial to a well-functioning system of laws and institutions. A leader, in this sense, is someone who has both the authority and responsibility to fill a gap in the law by setting an official course of action or taking an official position that will directly impact the rights, welfare, or interests of citizens in the present and future and who must do so in a context involving some degree of risk and limited information.

Section I of this paper will offer a breakdown of government activity as a whole in order to describe what I mean by gap-filling activity and how it relates to other types of government activity. Sections II and III will use this breakdown to draw out existing guidelines for discretionary reasoning in scholarship dealing with the legislative and judicial branches. Detailed attention will be given in Section III to scholarship regarding the nature and proper use of discretion in the judicial context because the central question of this paper has to do with the proper use of discretionary authority. It will also further explain why executive branch gap-filling activity at lower levels, plays a role in the formation of the body of law in general that is just as important as the other branches or that branch's highest levels. Section IV will explain what principles may be applied to activity in the executive branch by analogy from the legislative and judicial contexts, and what those principles tell us about how executive decision making and executive discretion fit into a larger framework for decision making. Section V will use the example of standardized decision-making procedures used

by the U.S. Army to show how the unique demands of executive decision making shape priorities—and therefore standards—for decision making—in that branch specifically.

Section I: Authority Vacuums and the Leadership Branch

As a citizen of a modern representative democracy with a large and complex military and a vast administrative state, there are a few important propositions about the legitimate use and flow of government authority that I am disposed to accept as a matter of course for minimal requirements of acceptable government decision making. Because the exertion of government authority over me or on my behalf, even when it is beneficial, is unavoidably a usurpation of some part of my personal and political autonomy it must be both legitimate and justified. My autonomy is very important to me, and I should not accept its abridgment lightly—even if I am so accustomed to living this way that I rarely even notice it. When I *do* notice, either because I have taken the time to consider my situation carefully or it has thrust itself upon me in some way that I can't ignore, I am entitled to demand that it has not done so without proper justification.

One of the very basic propositions I believe about my government (and others) is that it ought to conform to some basic notion of a rule of law. This means that I have an expectation that government authority, when it is used, will be used only by people to whom it has been appropriately delegated in accordance with the law. Proper delegation alone, however, is far from enough. Any particular action a government takes, or decision it makes, ought to be appropriately justified for its own sake—there should be articulable reasons and they should be reasons of the right kind. That authority was properly delegated to the decision-maker matters, but it is a separate question. So, when I consider the acceptability of government authority as a force capable of overruling or superseding my autonomy in numerous ways, I find myself asking in the first case whether or not the decision was made in an appropriate way and whether it was made for appropriate reasons. In the second case, I am inclined to ask whether authority throughout the

government is distributed and exercised in accordance with the rule of law (such that institutions, procedures, and officials are all equally subject to the law) and some set of democratic norms that include things like transparency, accountability, and popular participation—or at least representation. What am I to think, then, when I discover that vast portions of my government's authority have been distributed to hundreds of thousands of nameless, faceless un-elected bureaucrats and military leaders as vague grants of discretion to execute or administer the law? For ordinary citizens, the government actions that most directly impact our lives are probably mundane, seemingly insignificant decisions made by such personnel.

If I am concerned enough, I might begin my inquiry by submitting extensive requests for information about any of their activities that might have impacted me under the Freedom of Information Act. During the long wait for a response that I am likely to endure as various agencies, departments, bureaus, and offices struggle to process my requests along with hundreds like them, I hope that it will also occur to me to demand a coherent account of what standards for reasoning these officials are expected to uphold while they make decisions about my rights and interests, and why they ought to be permitted to do so. Suppose I turn to published judicial precedent in cases where other citizens before me have been disturbed by a realization like mine. I may be disappointed that *Chevron, U.S.A., Inc. v. NRDC*,[18] the most influential legal precedent addressing an administrative agency's discretion to interpret and apply its own regulations—the best-known check on that discretion—requires only that they not be "arbitrary and capricious." That standard was promulgated in the context of the Supreme Court's declination to substitute its own evaluation of an agency's reasoning for that agency's when the agency acts in its area of expertise. Even the most astute justices might find themselves poorly equipped to interpret, for example, exactly what level of a particular contaminant regulated by the EPA is a safe level for industrial output, and what is

[18] 467 U.S. 837 (1984).

unsafe, such that they would prefer to require only that the agency show *some* rationale was employed to set the acceptable level rather than offer an opinion as to what that rationale ought to have been.

However, *Chevron's* requirements may still seem lacking to me if I compare it to the attention paid to whether and how the justices who established it employ their own discretion in cases that may be far less likely to have discernible impacts on my life in particular. When I consider the influence of government authority and the law on my daily life, in the Federal context, I am much more likely to think of the I.R.S. Regulations that control how I describe each portion of my income and expenses on my tax return than the long-ago act of legislation establishing Federal income tax in the first place. I am likewise more likely to think of whether I can visit a nearby national park any day I like, or I must choose a day of the week according to my license plate number or make a reservation weeks in advance. Yet, in both of these mundane cases, the requirements and restrictions that touch my daily life as an ordinary citizen are not strictly laws at all. They are creatures of policy and regulation. The same holds true in far more serious matters well.

A. More than a Machine: The Importance of Norms for Executive Reasoning

The executive agencies of the United States Federal Government operate nominally under the direction of *the* Executive—the President, or in the case of the military Departments, the same person under designation as the Commander in Chief. They receive their authority to regulate numerous facets of American life under various statutory mandates issued by Congress, which are at least to an extent subject to judicial review. Such a surface-level description situates them neatly within their place according to the national Constitution, making the extensive influence they can exert over private citizens—and non-citizens—palatable as a structure clearly and unquestionably consistent with the rule of law within at least a vaguely democratic system of government. However, they must regularly exercise government authority to make decisions that are characterized by high

stakes, urgency, and finality. They act in areas of uncertainty that open vast spaces between the bounds of legislative oversight and judicial review or even various mechanisms of branch-internal oversight. These decisions have significant and lasting—often permanent—impacts in the face of epistemic limitations, complexity, obscurity and lack of public interest or deep public disagreement.[19] In these cases, officials are called upon to either identify reasons for decision that others do not have the interest or ability to recognize (through the application of expertise) or to supply them in some non-arbitrary way where they do not already exist (through expertise, an act of leadership requiring the assumption of responsibility for risk, or some combination of the two).

To compound matters, the impact of these decisions is often immediate, and irreversible, and may be made in a context where either the demand for expertise or the requirements of national security limit or obviate opportunity for public oversight and accountability. For example, although there are avenues of remedy and discipline for a wrongly-made decision to launch a missile, the action itself is by nature final. Review of that decision is a qualitatively different proposition than the appeal and reversal of a judicial decision,[20] or judicial review of an allegedly unconstitutional statute. Likewise, a choice to deny political asylum, or to separate a child from her family, has effects that cannot be reversed even if the decision eventually is. Furthermore, decisive power in many of these cases is concentrated in individual officials who operate under only vague statutory guidance. This means that the authority of an ostensibly democratic government has its most direct, potent impact on individual lives through decisions made with little or no direct oversight by any elected official.[21]

---

[19] I have elsewhere referred to these cases as those where nobody knows, those where nobody cares, and those where nobody can agree.

[20] Even in the case of capital punishment, a lengthy appellate process separates the sentencing decision from the moment of execution.

[21] In some cases, there may be time to appeal to the judiciary for injunctive relief before action is taken or to seek some remedy through judicial review after the fact (as was attempted in *Locke*). Still, these avenues of oversight and recourse are limited. Likewise, Congress may call an agency representative to appear before it and give an account of how that

It is understandable that many people view the executive branch below the level of the President and Cabinet officials—especially the administrative state—as mechanical in its operation given the way that red tape and rote rule-following seem to characterize bureaucracy and strict obedience to orders the military. However, in even the most routine case of agency rule-making someone must approve the final draft of that rule, and its provisions once properly promulgated may be binding even if they turn out to be less than ideal (such as in the case of an allegedly confusing deadline for the submission of annual paperwork related to mining claims "prior to December 31," instead of the more customary "*by* December 31" or "*no later than* December 31," at issue in *U.S. v. Locke*, 471 U.S. 84). The decision of when the filing deadline ought to be was a gap in the statutory law left intentionally to the discretion—for better or worse—of whichever agency official approved the final draft of that regulation. Some such gaps are bigger than others. For example, an official working for the F.D.A. as a scientific expert may be called upon to determine whether available data indicates that a drug is safe enough for the market (making it available to improve or even save lives as quickly as possible), that the risk of harmful side effects is too great, or that too little data is available to make a reliable determination.[22] Likewise, a military commander at the tactical (lowest) levels of command operates subject to layer upon layer of higher-level guidance that controls what operations will be conducted, in pursuit of what strategic ends, for what larger purpose. Nonetheless, her position exists specifically because some individual at that level is required to fill in the countless remaining blanks of who will accomplish a mission and how. Those are decisions that may have far-reaching effects on a larger framework of operations and strategy, but

agency has wielded the authority it was granted, but a change in that grant requires a change in statute—a slow and often imprecise process. Even within the executive branch itself, control over an agency by the President may be limited, especially in the case of the independent agencies. Independent agencies, created by statute, are nominally part of the executive branch. However, they are not part of the Cabinet, and the President does not have the authority to remove their heads. These include the CIA, the EPA, the FTC, and the Social Security Administration (SSA).

[22]The case of FDA researcher Dr. Frances Kelsey's role in the disapproval of the highly-teratogenic drug Thalidomide is a famous example of the latter. (Epstein, 2018, 106.)

they may also have extreme and irreversible impacts on the individual friendly and enemy personnel involved in the conflict, or even civilians nearby.

Often, because such decisions tend to focus on future interests and future outcomes (and they often also involve the inherent unpredictability of a freely-acting human element, as I will discuss in more detail in the military planning context in a later section) there is no possible way that even a "perfect" reasoner could reliably identify pre-existing reasons that one course of action is definitively more rational, or better, than others. After all, if there was one clear way that the gap in question ought to be filled, the legislature or judiciary would most likely have already so filled it.

B. All Well and Good: Evaluating Decision Making

The questions I described above that I might have about the ways my government exercises its authority in abridgment of my own autonomy are questions about when, why, and how its doing so may be justified. To make this determination coherently and systematically requires standards by which these uses of authority may be evaluated. Setting such standards, in turn, requires us to first consider exactly what we are evaluating, or what those standards apply to. They may be intertwined but a decision itself—meaning the resolution or course of action decided upon—is a different thing from the results *of* that decision, and both are also distinct from the process of reasoning by which that decision was made. The first two have to do primarily with whether the decision is substantively good. The third has to do with the procedural question of whether it was well decided.

Both of these metrics appear in discussions of proper representation by legislative representatives, or proper use of judicial discretion by judicial officials just as they will both play a role in describing proper use of discretionary authority by executive branch leaders. In fact, it turns out as a whole that the criteria by which decisions are evaluated in representative and judicial contexts, as well as those applicable to leadership in the executive branch are the same set. Which of these takes priority, and the reasons why—in terms of the goals and purposes of the legal and

institutional framework within which an individual is entrusted with the discretion to make them—inform the appropriate standards by which those decisions ought to be evaluated to determine whether they are on the whole proper, or justified. The next section will explain in more detail what features activity in the three branches share in common, and how those features manifest differently in each according to their roles in the larger structure of government.

C. The Executive's Unique Gap-Filling Role

"The activity of the executive branch in the United States Federal Government" is so broad a category of activity that it would be impossible to say anything usefully specific about it as a monolith, and large portions of those activities resemble legislative or judicial activity. However, consider government—used as both a noun and a verb—as a series of interactions or set of relationships between laws and institutional structures established by laws on the one hand, and the officials making decisions within those institutions, in relation to those laws and structures, on the other.[23] Government activities can be divided into four general categories according to the operable relationship between that activity and the law: either making, interpreting, or applying the law, or performing some other sort of activity that is characterized by discretion. The distinction between one category and another has to do with the mode in which a decision maker relates, in a given case, to established law (in the broad sense of the term) and the discretionary spaces left open within it.

Of those categories, three are defined by straightforward modes in their relationship to law (making, applying, or interpreting). The fourth seems characterized more by the *laxity* of that relationship—and the need for or grant of discretion. Every branch engages at times in every type of activity in at least a limited manner, but there is one sort most characteristic of each which is unsurprising given their creation to serve distinct purposes. The first category of "making" type

[23] The further set of relationships between citizens and these laws, institutions, and individual officials who make up government are beyond the scope of this paper, other than the extent to which the different demands of these relationships in different contexts shape the applicable standards for decision making.

activities are those where law is being made through the creation of statutory or regulatory rules, primarily associated with the legislative branch but also present by explicit design in the administrative rule-making process.[24] The second category of "interpreting" activities include those where existing law or rules are being interpreted: primarily the role (and the *raison d'être*) of the judicial branch, but present in the executive as well by design such as in the creation of the Administrative Law Judge under the APA.[25] That would appear to leave the third, and most straightforward category of activity, characterized by applying laws and rules, as a candidate to be the defining activity of the executive branch.[26] The mode of applying does seem like a natural fit for the arm of a tripartite government intended to "execute the laws." However, this is far too mechanical an understanding that leaves aside the truly important activity in the branch (popular aspersions

[24] Most of this activity is subject to the Administrative Procedure Act, codified at 5. U.S.C. § 551, et. seq., which sets out the informal rule-making process typically called "notice and comment" rule-making. For many civilian regulatory agencies—for example, the FCC, FAA, or EPA, this activity is the most important—and publicly visible—of an agency's functions—the need to create rules to fill in the broad expanses of authority delegated through congressional mandate being the main reason for their existence. However, even those agencies whose most important functions are non-regulatory (such as those making up the Department of Defense and the other national security agencies, among others) engage at times in mundane rule-making or other typically administrative activities of the type subject to the APA, and not falling within one of its exemptions. (5. U.S.C. § 552 and 553, Folk 1986, Bonfield 1972). In these cases, agencies make rules that become part of a web of government authority we might call "the law" understood broadly, even if they are not statutory laws. This process is subject to only limited external oversight. Agency action taken under the APA is reviewable by the courts only when that action can be deemed "final," a standard not nearly so straightforward as an ordinary reading might suggest. At the time of this writing, the doctrine of Chevron deference (which, since it was established in 1984 in Chevron, U.S.A., Inc. V. NRDC, Inc., has instructed courts to defer to agency expertise on matters within their Congressional mandates) continues to ensure that a court will invalidate such action only if it was "arbitrary and capricious." "Arbitrary and capricious" generally means that an agency has acted without any reasoned consideration at all—with no comment as to the content or quality of the reasoning in question beyond the fact that its existence can be demonstrated. Furthermore, agency actions are subject to judicial review under these provisions only when they are "final," a distinction that is itself vague and subject to much-litigated interpretation. (*See, e.g.,* Hayes 2017, *Soundboard v. FTC*, *USACE v. Hawkes*, *Bennet v. Spear*).

[25] In the executive branch, these activities also include less direct analogs like the activities of the National Labor Relations Review Board (an independent agency) and other boards with adjudicative powers, various internal investigatory functions throughout the agencies, including establishing insulated Inspector General offices within organizations, and internal "non-judicial" adjudication processes. In the military, for example, any commissioned officer, many non-commissioned officers, and civilian personnel may be temporarily appointed to adjudicate certain internal matters alone or as part of a board.

[26] In these cases, agencies may operate under both the legislative mandate of Congress or an Executive Order and the regulatory oversight of another agency. These include adherence to tax codes promulgated by the Internal Revenue Service (IRS), the staggering bureaucratic process of Federal personnel management by the Office of Personnel Management (OPM), and the byzantine expenditure and acquisition strictures filtered through Constitutional declaration to statutory budgeting authority, and the ever-reliable Congressional budgeting process to the Office of Management and Budget (OMB), variously into the details of individual agency appropriations and the Federal Acquisition Regulation (FAR)… among other authorities.

about slow-moving bureaucratic machinery aside). First, this is easy to see in Presidential powers like the executive order, veto, and role of the Commander in Chief. It is just as true *and* just as influential throughout lower layers of authority where even broadly "insignificant" decisions such as whether or how heavily to redact a set of documents released under FOIA, whether to state an application deadline as "before" or "after" December 31st,[27] whether a soldier's request for conscientious objector status ought to be granted, or whether or not to bring federal criminal charges—or which charges to recommend and how to phrase them—against a petty offender who just happened to cross state lines and the decision whether to arrest that same offender in the first place. Second, it would be hard to make the case that the simple application of laws and rules could be particularly distinctive of any one branch. That government activity is carried out in accordance with (that is, through the application of) existing laws and rules across all branches, agencies, departments, and so on is an indicator of the rule of law and little else.

This leaves me the option of either concluding that there simply isn't anything distinctive about executive decision making or locating it in the fourth category of activities where decision-makers are doing "something else," other than making, interpreting, or applying law. In these cases, an official uses their discretion to act where the law is otherwise vague or silent. It is not clear though, that this means they are not in some degree still making, interpreting, or applying *some* part of the law, but what is clear is that none of those are in such cases the dominant mode of activity. For example, consider the location of law enforcement agencies (including the Department of Justice) within the executive branch. When an official must make a choice about how or whether to *enforce* a law that decision lies at an intersection between simple application and interpretation but is strictly neither. Instead, it is most significantly an exercise of individual discretion. The rejection of

[27] *United States v. Locke*, 471 U.S. 84 (1985).

an application for a mining claim submitted on December 31st because the controlling regulation set a deadline "before December 31st" at issue in *United States v. Locke*, is similar.

One way to describe this would be as follows: each branch may engage in a variety of activities spread across the first three categories (making, interpreting, or applying law). The legislative branch is primarily dominated by a mode of making law so that the bulk of its activity falls into category one, and the judicial branch is dominated by a mode of interpreting the law so that the bulk of its activity falls into category three. The mode of *applying* the law (and therefore category two activity) is necessarily pervasive in all of them, and the activities of the executive branch are so varied that it is dominated by none in particular, though it often performs activities that combine more than one.

This characterization is inadequate for at least two reasons. First, it fails to give an account of what particular role the branch plays, such that its existence distinct from the other two branches is necessary for the work of governing as an activity fundamentally related to the law. Second, it does not describe decisions made by agency experts and leaders, especially military leaders, other than that they are vaguely instances of "applying" or "interpreting" laws that have granted decision-making authority to those leaders.

The fourth category of activity is a category in its own right, not just a catchall for combinations of activities belonging primarily to the legislative or judicial bodies with rote application. Understanding why calls for clarity about what constitutes activities "making" law. There is a sense in which *every* act of governance makes law to the extent that we think of the overall body of the law as including things like regulation, judicial precedent, and executive orders. They are all just different types of decisions made by some official or other about how government authority ought to be used to order the lives and protect the interests of citizens. This would collapse the other categories into category one in a way that is just as inadequate as the reduction of category

four to a mere modifier. A narrower sense of the law made in category one that is restricted to making written rules according to well-established procedures avoids this collapse. Cases in which law is effectively made by an executive branch official or a judge because there is nothing decisive to interpret or apply remain, circling back to category four as the home of discretion. What description might fit the mode of relationship to the law at work in activities where the law in question is only a vague grant of discretion or legislative (and judicial) silence—in some sense, the absence of law?

As I stated above, it would be contrary to the basic requirements of legitimacy to create a vast branch of government imbued with extensive discretionary power without some justifying necessity.[28] Clearly, the executive branch does serve a distinct purpose. There are decisions made within it that could not possibly be handled by popular vote, legislation, or carefully considered judicial decision. They call for expertise, quick leadership, or both, and they are made by the authority of discretion granted where either design or the inevitability of chance have left gaps in the law. In doing so, they are not making, applying, or interpreting laws, but neither are they suddenly operating outside of *any* relationship to the law. They act with discretion that they are granted by law to add something to it. The mode that defines the executive branch, therefore, is a mode of gap-filling. Just like the branch itself combines activities that mimic the others, its defining mode combines aspects of making, applying, and interpreting law with expertise and leadership. The element of expertise speaks to a defining purpose in many of the administrative agencies and the branch-specific priority for epistemic fulsomeness and accuracy. The element of leadership speaks to the way that executive decision making is more likely to be results-focused than judicial decision making is, and consistently more likely to call for the employment of internally supplied reasons

[28] Identifying that necessity is important to the extent that we are concerned with respecting necessity as an important norm for legitimacy, by which I mean the idea that we ought not to govern—by either imposing external power to order the lives of citizens or to act in their names—without need. Doing so is unavoidably an infringement of personal autonomy. I do not mean to take an anarchist position by highlighting its importance. I do not contend that it cannot be justified, but that it must be.

than either of the other branches. Most of these decisions are made by one person, who—because of the reliance on her own internal judgment and the often immediate, irreversible impact of her decisions—must bear the risks and responsibilities for the results of her decision in a direct, personal way.

However, that does not mean that officials in the executive branch are *never* called to make decisions that must place a greater priority on procedure, or that officials outside of the executive branch *never* exercise this type of individual, decisive leadership—only that it is comparatively rarer. The mode in which government officials act is a product of their relationship with the law. The relationship an official has with the law in each case is shaped not only by the case itself but also by the purpose within the institutional structure of government that a particular official's branch primarily serves. Multiple layers of *what* must be decided and *why* it must be decided influence the requirements for "how." To return to the broad distinction between well-decided and good drawn from scholarship focused on the legislative and judicial branches, I can make some general claims about how that priority shifts. The balance between the shifting sets of priorities discussed in section three (criteria for evaluation, goal to be served, and method of selection) is responsive both to the overall institutional role of the decision-maker and the part of government to which she belongs and also according to which mode of relationship with the law or category of government actions is in play. To the extent that anything other than straightforward application is taking place. For example, an "interpreting" type activity in the judicial branch will always be the most characterized by concern for consistency with past decisions, as compared to a "making" type activity in that branch or an "interpreting" type activity in the executive. This is why the reasoning we would expect to see in each branch will be different according to its overall role in government—the institutional, as opposed to individual, purpose for which the decision is made—as well as the specific characteristics of the decision itself.

Legislators make laws according to procedures delineated loosely in the Constitution and elsewhere in great detail—to the extent that there are roles like that of Senate Parliamentarian dedicated entirely to ensuring that correct legislative process is followed[29] and there is much scholarship focused on what constitutes proper and effective representation. Judges likewise decide cases according to rules (existing statutes) and decisions (the body of precedent), and volumes have been devoted to attempts to describe the reasoning process by which they do so, or ought to. In the case of executive branch officials, similar guideposts exist in some areas of their activity but their application is specific and therefore limited. Statutes and regulations like the APA's notice and comment procedures and the FAR's requirements for soliciting, evaluating, and selecting contract bids. Other statutes and regulations prescribe procedures for hearings during formal APA rule-making or military investigations.[30] However, none of these attempts to hem in discretion with procedural guidance speaks in the same sort of general terms that theories of representation or judicial reasoning do or identify broadly applicable standards for gap-filling discretionary decision making similar to those that we have for legislative and judicial procedures.

Although it would be a mistake to assume that norms for similar activity in other branches will apply directly to even those instances in which executive branch officials make, interpret, or apply law in the same way that they do to officials in other branches, let alone to gap-filling activity, there are shared norms for government activity in general that underpin those standards and apply across branches even if they may be weighted differently. As the next step towards identifying some, sections II and III will discuss standards of evaluation for activities in the legislative and judicial branches, in turn, to identify which features may be analogized to the executive.

---

[29] *See, e.g., Rules of the U.S. Senate*, at https://www.rules.senate.gov/rules-of-the-senate, and House and Senate Rules of Procedure: A Comparison, CRS Report 30945, 2018.

[30] For example, Army Regulation (AR) 15-6, Procedures for Administrative Investigations and Boards of Officers is a 68-page document covering everything from rules of evidence to how investigating officers should handle information subject to the Privacy Act.

Section II: Analogies to the Legislature and Theories of Representation

In this section, I will discuss theories of representation pertaining to the activity of the elected representatives in the legislative branch. These theories provide three valuable insights for describing the reasoning process that executive decision-makers ought to engage in or identifying the norms by which that process ought to be evaluated. First, the conversation surrounding how methods of representation can be described or should be evaluated (as a normative matter) makes clear that different methods of representation are descriptively applicable (and/or normatively appropriate) for different government functions. Second, the norms by which representative activity ought to be evaluated as normatively correct or not are different according to which of these models applies (and perhaps, according to which sort of government function is being carried out). Third, one important underlying distinction among methods is whether or when an official has the authority to supply such reasons himself and when he should adopt them from an external source.

I am focused on theories of representation, as opposed to the details of the legislative process, because it is only *as* a representative that a member of a legislative body in the U.S. or any other representative democracy has the authority to participate in the making of law. Any decision such an official makes is always made as, and by virtue of their being, a representative.[31] Therefore, any decision such an official makes is always rooted in a choice about how, who, and what they are to represent and whose judgment ought to be deferred to in the hard cases—their own, that of the constituents who expressed preferences by electing them, that of some segment of the population with whom they identify, and so on. Any reasoned decision process in that branch is necessarily rooted in decisions about representation. This means that the principles of description or normative

[31] To the extent that the President (or a primary Executive official in general) is also an elected *representative*, such theories would, of course, pertain directly in a way they do not to the non-elected officials that are the focus of this paper.

evaluation about reasoning in the legislative branch that might apply in the context of the executive activity this paper is focused on will most likely have to do with theories about representation.

Pitkin (1967) sets the backdrop for contemporary discussion on the subject, engaging with the traditional divide between views that representatives should (or do) act as either delegates selected to act in accordance with the judgment, opinions, and desires of their constituents or trustees empowered with the discretion to exercise their own judgment about what is in the interests of their constituency.[32] The key points she makes relevant to this discussion are that neither of these models nor any others proposed[33] successfully capture the whole of what political representation either is or ought to be, and that multiple theories have their appropriate uses in different contexts. (227). According to Pitkin, the common yardstick to be applied is whether a government as a whole (as opposed to any particular act of any particular representative) is or is not representative, understood as a matter of whether it is demonstrably under the control of its citizens instead of the reverse, which requires a combination of responsiveness and accountability. (231-2).

More recent work in the field has sought to further refine or expand on the notion that representative activity is not always properly either delegate/responsive or trustee/fiduciary activity alone (or any other single type that has been identified), but that instead that there are different models of representation that may appropriately serve different norms in different situations.[34] Jane

---

[32] *See*, *e.g.*, 119-120, discussing analogies for representation; 191-2, discussing Edmund Burke and James Madison's differing uses of the term "interests".

[33] Pitkin identified four broad types of representation that we *might* mean when we invoke the term, each encompassing its own sorts of representational activity and subject to its own standards of evaluation: formalistic representation (having to do with the institutional position of the representative, whether as a matter of authorization or accountability), symbolic representation (having to do with the representative's acceptance by the public *as* representing them), descriptive (having to do with the degree to which a representative shares characteristics in common with those represented), and substantive (having to do with the substance of the representational activity itself). (1967).

[34] Mansbridge identifies the prevailing tradition in representational theories, both preceding and following the time of Pitkin's work, as "promissory." Both sides of the trustee/delegate debate can be seen as versions of a promissory theory. Promissory theories of representation deal with promises made during elections that representatives either keep or do not, for which the appropriate normative criteria is "democratic accountability." (515). Mansbridge identifies three other legitimate forms of representation that occur: anticipatory (representatives acting out of concern for their chances of reelection in response to their beliefs about what the wishes of the public *will* be, not what they have been previously expressed to be), gyroscopic (models of representation where a representative is selected because voters expect her to

Mansbridge (2003) argues that we should consider not whether a representative is acting in one acceptable way but whether they have satisfied the norms that apply to whichever model (or models) of representation apply to the activity in question (526). Andrew Rehfeld (2009), argues in a response to Mansbridge that the historical focus on the trustee/delegate debate is the result of a mistake about the operable question in that debate.[35] It is not because representatives represent that we are concerned with identifying the proper location of authority to decide (with the representative as a fiduciary or with those represented, for whom she is a mere delegate), "but on account of the role that a representative takes on as a decision maker." (229).

Representatives are just one type of decision-maker among many who must identify or supply reasons for decision, whether those reasons are the product of the representative's own judgment—internal reasons—or the judgment of the people—external reasons, accepted in the traditional sense of "responsiveness." Setting aside the question of which taxonomy of representational methods, if any, is "correct," the principles that 1) there are multiple legitimate methods for decision, 2) which method is appropriate depends on the type of decision to be made and the context it will be made in, and 3) a major distinction among methods that remains consistent throughout the debate is whether the decision-maker is empowered (or expected, or required) to supply reasons of her own, or she must adopt reasons that are already available and identifiable.

---

act in certain ways that they desire "for her own internal reasons," (520), because they believe that those actions will further their own interests), and surrogate (which occurs when representatives represent "constituents outside their own districts," meaning citizens *other* than those that they are functionally accountable to). These models, Mansbridge says, are all legitimate in their own ways despite their failure to exhibit the same sort of democratic accountability that features in the promissory model. (526).

[35] According to Rehfeld, focus on the tension between responsiveness and independent representative judgment as a question about representation (whose judgment ought to be deferred to) obscures principles that describe the obligations of any decision-maker whether they are a representative or not. The trustee/delegate debate asks not about who is entitled to decide ("location of authority"), but instead is a more multi-faceted question about 1) the proper aims of representation (meaning whose good the decision should be made for), 2) the proper source of judgment about the substance of that good, and 3) how responsiveness to the desires of those whose good is aimed at should be accounted for. This reframing of the debate has the advantage, Rehfeld says, of separating the important question about the proper location of decision-making authority in *any* instance where law is to be made from the issue of representation. (216-17). The proper location of authority, Rehfeld says, is an entirely separate question from the three encompassed by the trustee/delegate debate. (217).

These principles track the division of government activity in general and executive branch activity in particular into multiple categories of activity, each dominated by different modes for relating to the law as outlined in Section I of this paper. The difference is that in the discussion about political representation, the question concerns the relationship a representative, who is specifically granted the authority to make law *because* the public has elected her, should have with the public. In this paper, the question concerns the relationship a decision-maker in the executive branch should have with the law and with the public in carrying out the various functions those same elected representatives, and the laws they enact, have delegated to her. In any case of government decision making, the decision-maker ought to act on the basis of clearly articulable reasons. The differences at work have to do with where those reasons should come from.

Theories of judicial reasoning have much to tell us about when and why a judicial decision-maker ought to rely on what I have described above as internal versus external reasons. In the next section, I will explore that discussion to tease out further common principles of government decision making that apply across branches.

Section III: Analogies from Judicial Reasoning

In the last section, I discussed the debate regarding theories of representation between those prioritizing internal and external sources of judgment, as well as more recent literature arguing that different modes of representation, in which representatives seek reasons for decision from different sources and according to different priorities, are appropriate in different contexts. This position parallels my claim that government officials also operate in different modes of relationship to the law in different contexts. This means that we are now dealing on one side with a choice between modes for relating to the law according to which category of activity an official engages in (making, interpreting, applying, or discretionary gap-filling) and on the other different modes of judgment according to where reasons for decision should be sought (internally, such as according to an

official's independent discretion or externally, such as according to an available set of facts or expressed public preferences). In this section I will explore principles from theories of judicial reasoning that discuss how a specific set of government decision-makers (judges) bridge the divide between these two sets of choices to determine whether or when discretion may be applied. That discussion of judicial reasoning offers us a detailed look at what sorts of reason-selection methods and what sorts of reasons are acceptable and under what conditions. The nuances of when and why a judicial decision-maker may be justified in employing discretion will provide a framework that can be applied for balancing any government decision-maker's internal judgment against external sources of reason across branches. That framework will ultimately provide a vocabulary to explain how the balance should be struck differently in the executive branch where constraints, purposes, and priorities are different.

A justified judicial decision, like any justified decision, is one made 1) based on reasons, so long as 2) those reasons are either the right reasons, the right kind of reasons, or reasons that were selected in the right way. Which of the criteria listed in 2) applies will vary with the situation. Judges look for these reasons in codified legal rules such as the Constitution, statutory law, or applicable regulations or common law precedents. When none are available, they may turn to more general legal principles (Raz 1972, Dworkin 1967) or vague notions of policy or justice.[36] There debate over whether and when judicial discretion is called for and how judges should identify, or if necessary supply, the right kinds of reasons to justify their decisions in the absence of statutory authority illustrates the kind of mode-switching officials in all branches employ in their own ways.

I've said that decision making in the executive branch is distinct from that in the other branches because it is most characterized by discretionary gap-filling activity, as opposed to making,

[36] *See, e.g.,* Dworkin 1986 at 228-258, describing the judicial interpretation of precedent as an attempt to author an installment in a "chain novel," D'Amato 2011, arguing that judges interpret the law according to socially-informed notions of justice.

interpreting, or applying law. To provide standards for decisions made as part of that activity requires a concrete picture of what that gap-filling discretion is, what makes it consistent with democratic norms, and when it ought to be used. The best point of departure for that effort is the similar discussion about judicial discretion. Discretionary reasoning performed in the judicial context, for judicial purposes, according to the constraints and priorities of the judicial institutional position is still *discretionary* reasoning. The question is generally "what makes it proper for a government official to invoke discretion as opposed to relying on other authorities?" An answer specifically about judicial discretion, such as that "discretion is proper when no statute or common law precedent decides the case at issue" would not tell us much about how and when an executive branch gap-filler should proceed, but an answer such as "a government official ought to rely on a grant of discretionary authority when no applicable external source of reasons for decision is available" certainly will.

Theories differ as to what exactly judicial discretion is or whether it exists at all. Dworkin (1963), for example, takes issue with popular descriptions of judicial discretion as "choice" unconstrained by legal standards even where *rules* do not apply:

> If the judge were free to adopt his personal preferences as legal standards, then indeed, his decisions would be chosen. But he is not. He is subject to the overriding principle that good reasons for judicial decision must be public standards rather than private prejudice. (635).

There is also disagreement about whether common-law precedent should be treated as a source of additional rules for judges to identify or situations in which they may instead seek analogy or disanalogy that can guide them through the formulation of new rules through the use of judicial discretion. However, even the most liberal concept of judicial discretion does not go quite so far as to allow free adoption of "personal preference" alone. These theories have in common the

recognition that predictability and stability in the legal system[37] depend on the courts' treatment of precedent as binding in some way. Deference to applicable precedent is just as crucial to maintaining a recognizable rule of law, and the stability and predictability that go with it, as the subjection of government officials—courts included—to the codified law. It may be more so. Codified law is finite, and the gaps revealed by unforeseen and unforeseeable circumstances that discretion must fill are potentially infinite.

Scholarship dealing with how judges reason with and about common-law precedent would most obviously translate to any of the judicial functions of the agencies themselves when we might say they are purely interpreting—as opposed to simply applying or making law—such as the decisions of the National Labor Relations Review Board, or the immigration courts that are a part of the Department of Justice. However, these distinctions between categories of actions and whether an agency (or judge) is, in a particular instance, interpreting, applying, or making law are less a reflection of discrete activities isolated from one another and more descriptors of the spectrum of modes in which government bodies may interact with established law.

These modes are fluid. For example, a judge may apply clear statutory law to some issues in a case, interpret a vague statute or take guidance from past precedent to resolve others, and then still be left with other questions that cannot be resolved by either of these approaches (applying or interpreting). These remaining areas of uncertainty still require decisions to be made, and the judge must identify some rational basis for that decision, but the authority on which that decision rests is the judge's own discretion. However, such decisions made through the application of judicial discretion subsequently become part of the body of judicial precedent—the common law—which

[37] *See* Schauer, 1987, for a discussion of the divide as "rules versus reasons," referring to Dworkin's approach from *Law's Empire* as a "jurisprudence of reasons" (rather than rules). This dispute falls out of the larger debate between positivism and non-positivism discussed in Shapiro, 2007, as the "Hart-Dworkin Debate", and is the heart of Alexander and Sherwin, 2008; *see also* Dworkin, 1963; Schauer, 2008, 2009, 1991; Sunstein ,1993; Stevens, 2018, 4-5.

interprets and supplements the work of the legislature (and the rule-making agencies, in some instances) and is sometimes referred to as "judge-made law."

Likewise, officials in the executive branch issue rulings, create rules, and take important life-affecting actions that carry the effective force of law as properly authorized uses of government power.[38] A person whose interests are directly and materially affected by one of these decisions is likely to be more concerned about the fact that the government has circumscribed their autonomy for good or for ill than with whether it was an administrative decision instead of a statute or judicial ruling that has so altered their fortunes.

Acts of executive discretion, like judicial rulings and legislation, are an avenue through which generalized grants of government authority solidify into a web of binding political authority within which we live our lives, and which we might refer to collectively as "the law," though it includes much more than just the assortment of codified statutory laws. One way to look at this is by expanding our use of the term "law" in two steps. Statutory laws are just rules with special political status because 1) they have been made in the right ways, and 2) their having been made in these ways renders them enforceable or in some other significant way potent. The first step of expansion is to recognize that statutory laws are not the *only* such rules. They are joined by articulable rules of precedent and the results of agency rule-making, which should, therefore, also be included in our understanding of what counts as law. The second step is to recognize that in addition to such rules, the law in an expansive sense also includes principles[39] that may not be reducible to the formula of a

[38] Interestingly, Dworkin, 1963, at 633 rejected the idea that judicial decision-makers possess the same kind of discretion that (business) executives and military leaders do.

[39] Raz (1972) makes a case for this inclusion of principles that cannot be reduced to rules in the service of his distinction between norms and standards of law and other socially accepted norms and standards (Raz 1972). That paper was a response to Dworkin's rejection of such a distinction between legal and social norms. In Raz's description, Dworkin rejected the distinction wholesale. (Raz 1972, 823). In Dworkin's paper, however, the rejection comes specifically as a rejection of Hart's rule of recognition and similar positivist approaches. (Dworkin, 1967; *see also* Shapiro, 2007, 7-8). A rule of recognition is some established rule that distinguishes those standards having the special status of law from those merely enjoying general social acceptance. (Hart 2012). Raz ultimately concludes that Hart's criteria for

rule but which carry legal force nonetheless. These include guiding norms like "fairness" and "transparency."

In Dworkin's view, a judge deciding a "hard case" where no rules are available should be obligated to base her decision on standards and principles rooted in morality that find their way into the law by default rather than recognition. (1963). Even in the hardest cases, that obligation to decide in accordance with the moral standards underlying the laws, if nothing else should be available, would make the difference between the public reasons he endorsed and mere private preferences. (1963, 632, 638).[40] For Dworkin, rules have an "all or nothing" character—a rule either applies or it does not, and if one purported rule conflicts with another, then one of them must not be a rule at all (since rules do not admit of exceptions, nor can they be prioritized over another to resolve a conflict). (1967, 27). Principles are more flexible, having "a dimension of weight or importance." (1967, 27). On this view, a judge who takes herself to be exercising discretion to modify or distinguish some legal rule is either acting improperly or making a mistake about whether she is dealing with a principle instead. A principle is no less part of the law, but principles *can* conflict with one another (which, according to Dworkin, is the key difference between the two). What is often mistaken for discretion, as Dworkin describes it, is a judge's effort to determine the

identifying the limits of the law ought to be modified in that a rule of recognition alone is insufficient because it cannot account for the status of legal principles. (Raz 1972, 841-42).

[40] The position that things other than codified legal rules, and perhaps properly decided judicial precedent, are "part of the law" is one of the most divisive in legal theory, forming the fault line between legal positivists and natural law theorists and at the heart of what is referred to as the Hart/Dworkin debate. The treatment of that disagreement offers another window into more generally applicable notions of what not only judges, but government decision-makers in general, ought to do when facing a case that appears to call for the use of discretion. Shapiro (2007) describes the divide between Dworkin and Hart as whether morality is (automatically, inescapably) part of the law or not (the heart of the divide between positivism and natural law). For Dworkin, it very much is, which is why he (in Dworkin, 1963) prefers the term "judicial obligation" to "judicial discretion." Dworkin feared that Hart's reduction of law to socially and politically adopted standards which (though they *might* have sources in morality) gain compelling force only when formally recognized allow judicial discretion to import personal preference into law. (1963). Having rejected the positivist notion of judicial discretion as a means of filling in empty spaces left between recognized rules of law, Dworkin provided an alternative account of what, in his view, already existed in those spaces.

relative weight and importance among principles that she is no less obligated to apply than she would be a relevant rule.

Dworkin's rejection of the term discretion aside,[41] two important facts are apparent. First: there is *something* in the law that is sometimes vague and contradictory, judges are called upon to make this vagueness specific and resolve these conflicts, and the results of their doing so ultimately take on the binding force of law. Second: the reasons a judge invokes to justify her doing so must be somehow "public" in character, even if she has no source of external guidance to rely on in selecting them. Here a parallel becomes evident between what legal scholars have to say about how judicial discretion should be used, whether they're happy with that terminology or not, and what justices have had to say about how administrative agencies ought to use their discretion in the executive branch. Acting with discretion is not the same thing as acting in a way that serves private interests, or that is either arbitrary *or* capricious.

Raz (1972) continues the useful discussion, stating that "[t]he exercise of discretionary powers is typically guided by principles rather than rules" and then that "[s]ome areas, such as governmental activities, cannot be adequately regulated by rules, and they must be directly governed by principles." (1972, 841).[42] Previously, he had acknowledged in passing that "the activities of

---

[41] Later thinkers have sought to incorporate recognition for legal principles, in addition to rules, back into a theory founded on Hart's instead of conceding the point to the natural law contingent, showing that they are not necessarily incompatible with talk of discretion.

[42] Raz adopts Dworkin's inclusion of principles as a part of "the law" that is distinct from legal rules but rejects Dworkin's description of the distinction between them. Raz argues that rules can and do conflict, and they too may be entitled to different relative degrees of importance—overall or in a particular case. Part of the judge's work is to resolve these conflicts by identifying which must take priority, which "is a result of legal policy," not a fundamental feature of their nature. (Raz 1972, 834). For Raz, the difference between rule and principle is one of purpose rather than authority. Rule and principles alike state "reason[s] for action." (Raz 1972, 836). The difference lies in the action prescribed: "[r]ules prescribe relatively specific acts; principles prescribe highly unspecific actions." (Raz 1972, 836).

Having presented his adjustments to our understanding of what a legal principle is, Raz suggests that we should add to H.L.A. Hart's positivist description of a legal system as those rules encompassed by an overarching rule of recognition (as having authority) that it also includes

> Not only…one customary rule of the law enforcing agencies and all the laws recognized by [that customary rule of recognition], but… all the customary rules and principles of the law enforcing agencies and all the laws recognized by them. (1972, 851).

public officials and administrative agencies are largely governed by principle [as the sole ground for action in particular cases.]." Such a formulation underscores the importance of those "particular cases" in which public officials and administrative agencies must look to vague principles to find reasons for action. Not only do those officials and agencies conduct activities that have the effect of directly creating law, but they also include the formal structure of law enforcement per se, as well as many other cases where an agency decision amounts to a choice over which parts of the law will or will not be actively enforced and to what degree. Raz calls judicial reliance on principles instead of rules "exceptional," which would be consistent with the judiciary's important institutional role in maintaining stability in the law; the case of the executive branch is quite different.

Taking Raz's understanding of the relationship between legal principles and legal rules as a touchstone, we can abstract a series of balance points between dichotomies of competing priorities that are present in both judicial and executive branch decision making. These pairs track where, in a given case, priorities are to be placed as to 1) the criteria by which the decision should be evaluated, 2) the institutional goals the decision is to serve, and 3) whether the decision-maker is in a mode of rule-following (whether by application or interpretation), or rule-creating. These balance points are linked to whether a given decision can be made according to an available rule or a principle must be employed instead. That distinction is, in turn, linked to the legal, institutional, and practical context in which the decision is made.

A decision-maker's navigation through these balance points corresponds to shifts in the relationship between that decision-maker and some set of laws or rules. Those shifts are influenced by the degree of certainty available to that decision-maker (either about the facts surrounding the decision to be made, the outcome of one possible decision or another, or both). They are also influenced by the degree to which certainty (again, about either the context or outcomes), as opposed to some other desiderata, does or does not take priority. These related certainty concerns

cluster around the completeness and depth of information available (epistemic completeness), the accuracy of that information (epistemic accuracy), and predictability. They track choices along a spectrum that describes how decision-makers approach their search for reasons and where they are likely to search for them. That spectrum ranges from rule-applying, constrained, low- or no-discretion activities at one extreme and unfettered discretion in the absence of either guiding authority or some other sort of rational determination at the other. They call for decision-makers to balance conflicting priorities against one another in ways influenced by the surrounding legal, factual, and political context. They share the underlying distinctions between internally and externally supplied reasons for decision,[43] and methods for decision versus the content of decisions (substance versus procedure), with theories of representation from the legislative context.

The scheme of dichotomies that influence where a decision-maker should look for reasons, explained below, is as follows:

1) Criteria of evaluation (of decisions)
    A) Well decided
    B) Good
2) Goal to be served
    A) Consistency prioritized; stability as a goal
    B) Goals other than stability are prioritized (accuracy, efficiency, effectiveness, flexibility, etc.)
3) Method of selection (of reasons)
    A) Rule following (external)
    B) Rule creating (internal)

A. Criteria for Evaluation

The first balance to be struck is between the criteria used to evaluate a decision: whether it is A) well decided or B) good. In judicial reasoning, where Raz tells us the discretionary application of principles in favor of rules is "exceptional," the quality of being well-decided takes priority. The fact

[43] "Internally supplied" reasons should be distinguished from the "private prejudice" that Dworkin contrasts against public standards. Even in Burke's fiduciary model of representation, the representative is supposed to be using her private judgment about what is in the *public* interest, and she is empowered to do so because some segment of the public has specifically endorsed her judgment for that purpose.

that a rule exists, in this context, is more important than what its content might be. (*See*, e.g., Raz 1986). This is because the judicial system, at least in a common law system, must prioritize the stability of the law. Stable law makes for predictable law. Predictable law makes for a stable society because people are able to order their actions according to reliable expectations about whether or not they will find themselves hauled into court and what will happen to them if they do (or what will happen to their neighbor, should they need to see their neighbor hauled into court). A statement that people have rights under the law is of little value if they can't predict how those rights will be interpreted and enforced. If, on occasion, this means that a less-than-ideal precedent becomes entrenched in the law so that one bad decision must, by procedure, follow another, the longstanding hope is that the common law will eventually "work itself pure." (Dworkin, 1985). The type of certainty that takes priority in such decisions is institutional in orientation, and centered on the characteristic of predictability as opposed to epistemic completeness or accuracy: we need certainty about what the law, as an institution, is and about the procedures the judiciary will use to decide controversies about it.

In contrast, consider the FDA drug approval process. Certainty in the interest of predictability about what decision process the FDA as an institution uses matters, but certainty about the accuracy of 1) the information relied on in that process and 2) the corresponding quality of the agency's decision as a result of 1), is considerably more important if that agency is to serve the public interest well. It is unlikely that any responsible pharmacologist would suggest that instead of rigorously testing medications, we ought to introduce them immediately to the market and allow the available range of formulations to "work themselves safe," whether we could rely on the free market to favor safe medications over unsafe ones in the long run or not. Even the potential for litigation as to whether proper FDA procedures for such a decision had been followed would have been less important than the need for accuracy. What mattered, in that case, was the content and the

resulting outcome—ensuring that a drug would not be approved for the market if it could not be shown to be reasonably safe.

In the example of a judicial decision subject to established legal rules, there is greater epistemic certainty available (as to what the law requires and what the facts of the case presented are). There is also a greater need for predictability—because the decision occurs within a system that plays an important function in maintaining a stable rule of law. In the example of the FDA decision, there is less certainty of either of these types available—no rule can tell an official by-name which drugs ought to be approved and which ought not, and in that case, the necessary data was unavailable—and certainty in the guise of predictability, such as for drug makers as to whether their drugs will be approved, perhaps, or the public as to how long or how many studies approvals will take—is of little value. The values in play are better described in ways that sound like Raz's version of a principle: "Scientific officials in agencies ought to use their expertise to act in the interest of public safety."

B. Goal to be Served

The first balance point between whether we ought to prioritize the procedure or content of decisions—whether we care more for consistency or outcomes—links straightforwardly into the second. In a given case, determining what goal is most important requires deciding whether procedural (well-decided) or substantive (good) concerns must be prioritized. This choice is tied to how a decision-maker views her relationship to similar past cases. Again, a judge is much more limited here since the rule of law requires consistency. *If* applicable[44] past decisions exist, she is constrained by them because stability in the law is the goal. However, there are instances (Raz's "exceptional cases") where neither statutes nor rules of precedent decide all of the issues presented.

[44] I refer only to whether precedent exists and is applicable to avoid wading into the debate over rule-or analogy-based reasoning and the practice of distinguishing, or the related dispute over whether judicial discretion properly exists at all (*See* Alexander and Sherwin, 2008) which is beyond the scope of this discussion.

In the executive branch, it is far more often the case that either 1) there is no repository of past decisions available to be reasonably relied upon, or 2) even if there is, consistency for consistency's sake is either an unimportant or even unreasonable goal. For example, an agency directed by Congress to make a new rule addressing a matter previously the subject of another rule voided under the Congressional Review Act (forbidding the agency to make another rule "substantially similar" to the invalidated provision) cannot reasonably conclude that stability in the law should be valued above all else and that the proper course of action is, therefore, to decline the Congressional mandate that it now decide differently than the last set of officials tasked with writing such a rule. Assuming the Congressional mandate to make a new rule was vague—as they usually are—agency officials must identify some other goal to help them establish priorities.

This is one of the areas in which scholarship currently offers little concrete guidance—either for executive officials specifically or by analogy from the judicial. It is the realm of the vague guiding principle, whether we want to follow Dworkin in looking to morality or Raz in attempting to reconcile principles with Hart's positivist project. In either case, *some* reasoning process must be applied to ensure that whatever reasons are used to set those priorities are somehow answerable to Dworkin's "public standards."

C. Method of Selection (of Reasons)

We are then thrown towards the third balance point—to what extent the decision-maker will follow rules or must create them. In cases where reasons (applicable rules or precedent, factual certainty, or predictable results) are not readily apparent, the official must supply them herself. I have used the phrase "rule creating" here, but such a decision might invoke a principle instead so that it does not create a recognizable rule *per se*. It is also possible that a decision about which goal should be served itself supplies the necessary reason (because there is one particular choice that best

furthers that goal). I use the term "rule creating" anyway to highlight that these instances are the ones where discretion, in either branch, is at its most influential in forming the body of the law.

These three balance points describe different aspects of the same problem: what must be decided by an official granted discretion for what purpose (in support of what *institutional* goal?), and what facts (including facts about applicable statutes, regulations, rules of precedent, or legal principles) are available? In the order that I have listed them, they progress from general to specific layers of balance that will shape what sort of reasons a decision-maker should use to make a particular decision, but one set of priorities that I have placed on the list with the label "A" for each balance point lean towards consistency, continuity, and a focus on the process of decision making whereas all of the "B" alternatives lean towards flexibility and a focus on results. Which sort of approach is appropriate depends on the decision and the branch of government where it will be made. This is one aspect of the decision process that the second inquiry expresses. By asking what goals the decision-maker should prioritize, the inquiry requires that we consider both the goals of a particular decision and the framework of larger institutional (at the bureau, agency, department, and branch levels, for example) and government-wide goals. At each of those levels, the balance between the A-type and B-type priorities might differ, with the priority placed between them at each higher level influencing where the balance should be struck at the other points. A decision characterized predominately by "A" type features, where procedure is more important than results, and stability is an important goal to be served, is one that is unlikely to call for the application of discretion. A decision characterized by "B" type features, on the other hand, where results are more important than procedure and some other goal like efficiency is to be served is more likely to. A decision in which no satisfactory external reasons are available, in which case discretion may be the *only* available source of authority, is likewise not likely to arise in a context where institutional stability, secured

through consistency, is an important goal since the lack of external guidelines implies that there is nothing to be consistent *with*.

This is one reason for the differences between judicial and executive discretion. Where stability and predictability are important values of the judicial branch, the executive branch contributes (at times, at least) flexibility and agility. A judge cannot unilaterally overrule binding precedent (such as the decision of a higher court that it lacks the authority to overrule) or a properly enacted, constitutionally sound statute only because it has had disastrous substantive results. To permit otherwise would buy a few case-by-case victories at the expense of institutional stability—not just in the judiciary but the entire framework of the law. Yet we would not likewise wish for a military commander to repeat past mistakes in the name of consistency. We would consider that kind of consistency to have less than no public value because the institutional purpose and position of that commander *and her discretion* are fundamentally different, even if we might expect her to reason about her decision in a similar way by looking first for rules in statutory, regulatory, or military doctrinal guidance and then to prior similar cases for anything that might be useful guidance by analogy. Those past cases may have relevance to her decision, but they are not her primary focus, and she is unlikely to consider herself bound by them. Military doctrine is constantly evolving, as those institutions seek to learn from past successes and failures—because those decisions are made in a context where results are far more important than consistency.

As a matter of method, judges characteristically focus primarily on rules and looking to the past (barring opportunities to exercise true discretion, such as decisions made on the grounds of public policy). At their most characteristic in the category four activities involving expertise and leadership, executive decision-makers are concerned primarily with outcomes and looking to the future. In the final section of this paper, I will discuss existing decision procedures in the executive branch itself that attempt to tackle the problem of an orientation that, instead of drawing on a finite

body of past cases for guidance, must attempt to identify and consider some sufficient number of the most relevant *possible cases (outcomes)* in the future. First, however, there is more to be said about how the role and primary aims of each branch (such as justice, stability, and predictability in the law or the ability to exercise discretion with agility and flexibility) inform whether we are more likely to evaluate its decisions according to whether they are primarily A-type "well-decided" decisions or B-type "good" decisions.

Section IV: Applying Concepts Translated from the Representative and Judicial Contexts to the Executive Branch

This section will revisit the initial categorization of types of executive activity offered in Section I to address how those activities and their corresponding modes for relating to the law are influenced by Section III's series of balance points and the different roles and priorities of each branch to determine which sorts of reasons should be relied on, by which sorts of decision-makers, for which sorts of decisions.

From theories of representation within the legislative branch comes the general division between cases where it is appropriate for an official to act according to their own internal reason and those where external reasons are supplied or should be sought. From theories of judicial reasoning comes a series of balance points setting priorities for decision-makers that shift with different modes of relationship to the law. Underpinning both is a tension generated by the difference between whether a decision is "well-decided" or "good." Which metric and associated set of other priorities will dominate is, to an extent, a product of the type of decision to be made, but it is also influenced by which branch the decision maker is situated in. The question is tied to the role of each branch within the federal system and the corresponding differences in which democratic norms

take[45] priority in each against the background of a common concern for the overall maintenance of good governance and legitimacy. In every branch, a balance is necessary. However, that point differs between branches according to the purpose for which the authority to make certain kinds of decisions has been granted and within the branches according to which kind of decision is being made. These differences have to do with the different relationships to the law involved and how that relationship relates to the larger purposes of each branch.[46]

[45] For example, the legislative branch, since it is composed of representatives, is more beholden to be responsive to public opinion, for example, whereas the judiciary must be concerned with impartiality even where public opinion is strongly partial, while Constitutional guarantees of Equal Protection constrain both.

[46] For example, using the first balance point distinction between whether it is more important that a decision be procedurally well-decided or substantively good, the breakdown of these priorities by branch and institutional purpose is as follows:

A. Legislative branch, charged with making law as the representatives of the people: As Mansbridge discusses, appropriate approaches to decision making, understood as methods of representation, can vary. Which is chosen should generally consider whether the case is one in which process (such as responsiveness to popular opinion) or substance (such as a satisfactory result, even if it is obtained contrary to public opinion) is more important. The question of what "good" law is as to content is largely subjective, so although laws must be effective and lead to beneficial consequences of some kind, theories for evaluation focus on an understanding of good that is more analogous to being well-decided; they are made in the right way or for the right kind of reasons.

B. Judicial: charged with interpreting the law (and settling disputes) as impartial stewards. Well-decided takes priority in the interest of impartiality, predictability, and stability. Only in cases of silence in statute and precedent, or as specifically prescribed by law, and therefore arguably itself in obedience to prescribed procedures, may results (substantive goodness) be considered.

C. Executive: charged with executing the law—applying, enforcing, regulating, and generally acting (in accordance with the law) according to often-vague grants of statutory authority. Standards for "well-decided" are notoriously vague (e.g., the *Chevron* requirement that agencies merely not be arbitrary or capricious). "Good" also often takes priority due to the nature of the decisions made and the context they are made in.

2. Bias by category and mode

A. "Making" law (in the category one sense): situation dependent but warrants special consideration for public opinion and public interest. Non-legislative examples include the requirement for public notice and comment in APA rulemaking and the invocation of "public policy" grounds as a basis for judicial decisions made in the absence of controlling statute or precedent or as grounds for overturning past precedent.

B. "Applying" law (in the category two sense): straightforward rule-following. These decisions could be described as purely focused on process, and what is "well-decided" in that a decision-maker either correctly applies the law regardless of the results, or does not. If not, the action in question is either a violation of norms for government decision making, or (if it is justified and properly authorized) a departure into one of the other categories of action.

C. "Interpreting" law (in the category three sense): in accordance with Constitutional requirements, warrants special consideration for impartiality, which generally translates to an emphasis on decisions that are "well-decided" whether or not the results are preferred by either the decision maker, other officials, or the public. Appears throughout judicial and quasi-judicial procedures within the agencies.

D. "Filling gaps" in the law (in the category four sense): procedural constraints vague or lacking by nature (e.g., emergency decisions). Rare in other branches, except in cases requiring judicial discretion. Metric of well-decided is limited to the determination that discretion is 1) authorized and 2) required. "Good" decisions, generally meaning desirable results, are the priority.

Gap-filling decisions are overwhelmingly likely to be ones where the quality of the results (that they be substantively good decisions) takes priority. To illustrate why, briefly imagine the whole of a citizenry and their government institutions reduced to a group desiring to hold a barbecue. The government officials are now the group appointed to plan the event according to a written menu, agreed upon in advance (a stand-in for a constitution and bill of rights, since we will assume for the sake of discussion that even a barbecue is more enjoyable under the rule of law), that entitles the attendees (the citizens) to certain courses of food and beverage. Suppose that menu specifies only that there must be two protein courses available. A majority of the planning group was chosen specifically because they promised to ensure that one of those courses would be traditional ground beef hamburger patties, but there is no clear indication of preferences for the other. Because sausage of some sort is also a popular, traditional protein dish at barbecues, the planners agree relatively quickly that tradition is the next-best available clue as to what the attendees must wish to eat, so the second course should be "sausage." However, none of the planners are chefs or even passably competent home cooks, and the available guidance of popular opinion, menu mandate, and group tradition have all been exhausted. Therefore, they delegate power to a group of experienced culinary professionals to settle the details regarding what kind of sausage will be made, how they will be made, and to make them.

Those professionals propose several detailed recipes. They share the recipes with the barbecue attendees and solicit comments to determine whether anyone objects to any ingredients. An argument ensues as to whether two of the recipes—one for a common hot dog and one for a plant-based, vegan offering, properly count as "sausages" at all. Another group objects that the third recipe will be unacceptably high in saturated fat.

They all appeal to another person appointed as the recipe-interpreter and dispute-resolver, but that person, like the planners, is hopeless in the kitchen. His role is to settle disputes impartially

per the dictates of the planning committee—so long as those dictates comply with the written menu. In this case, they do. He concludes that the recipe writers have done what they were asked under the authority the planners gave them and that they were not arbitrary or capricious in their decisions. Beyond that, he can only defer to their greater expertise and the authority given to them by the planning body on the highly technical matter of what constitutes a sausage and what kind of sausage is best. By then, it is nearly four o'clock in the afternoon, there is only one hour left to purchase the necessary groceries, and there is only enough money for the ingredients to make one of the three recipes. Aside from the small minority who objected to one or the other of the recipes, the citizen attendees are only concerned with the fact that they are extremely hungry. If there is to be any barbecue at all, someone must decide which kind of sausages will be made and get on with making them (thereby filling the gaps left by the written menu, the direction from the planners to write recipes and make sausages, and the dispute-maker's declination to substitute his own lacking knowledge of sausage making).

To fully commit to the cliche: if we as citizens of complex modern governments are the barbecue attendees in this example, we should not indulge our desire not to know how the sausage is made even if it is true that we are unlikely to care about things like what procedure was used to select or write the recipe or chop the ingredients, or whether it was the same as the last barbecue. Absurdity notwithstanding, these sausages represent some of the most important, direct, and frequent impacts our government has on our lives and interests. Even if our priority is that the result be edible (at minimum), taste good (ideally), and be recognizably a sausage, if we are going to consume a lot of them, we should very much want to know certain things about how those sausages were made (what ingredients were used, at least, and whether the kitchen was kept clean). We can say several important things about what we expect from the people making them (and why) that go beyond asking only that they are not arbitrary and capricious. The fifth and final section of this

paper will discuss some areas of decision making in the executive branch that make clear what demands and risks shape decision making in its component parts when its officials must fill gaps in the law (without the results, ideally, tasting bad or causing food poisoning). Much like theories of representation for the legislative branch, and theories of judicial reasoning, some existing decision frameworks within the executive branch highlight these needs. The factors they are most attentive to illustrate some of the specific needs for desirable reasoning in that branch when its characteristic features of expertise or leadership are called for (much like responsiveness is often required of the legislature and concern for stability from the judicial) and clarify what that leadership is in the context of the gap-filling mode.

Section V: One Example of Standardized Gap-Filling in Action, and a Framework for General Principles of Leadership Reasoning

In earlier sections of this paper, I have presented a scheme of government activity, the relationship of decision-makers to laws according to the type of that activity, and the shifting balance points of institutional and contextual priorities that guide the use of discretionary power. In this section, I will discuss two prominent arenas of decision making in the executive branch to illustrate how these priority shifts manifest either when its officials act in capacities that mimic the other branches or when they act in a purely executive capacity and what they can tell us about the standards for decision making that specifically executive reasoning demands. The first sort of activity that I will discuss is the Administrative Procedure Act (APA) rulemaking, and the second is the U.S. Army's Military Decision Making Process (MDMP).

I previously claimed that executive decisions tend to prioritize results—substantive goodness—over procedural well-decidedness. One reason for this has to do with the branch's institutional role—specifically to "execute" or act—in the gaps left between the mandates of the legislature and the pronouncements of the courts. Though the goals they serve are varied, executive decision-makers act in capacities where they are granted authority specifically because there are cases

where efficiency, accuracy, sensitivity to individual situations, speed, or other goals requiring decisiveness and flexibility take priority ahead of stability for its own sake, responsiveness to public opinion, or even transparency. This institutional priority for good results forms a background that influences the appropriate balance points that guide decision making across the categories of decisions an official might make. This means that the bias will be at its most pronounced in category four gap-filling activities, calling for leadership and expertise, which also favor the good over the well-decided. This does not mean that those goals and norms more characteristic of the other branches do not matter, only that when decision-makers in the executive act in the ways that belong most to it, those goals are not the highest priority.

At first glance, rulemaking activity under the Administrative Procedure Act (APA) seems like the place to look for executive—and especially administrative state—activity at its most characteristic. Those procedures do give clear examples of where the important general norms underlying particular standards for reasoning in the other branches carry over into the executive—showing in various aspects measures to ensure transparency, responsiveness where possible and appropriate, stability, predictability, and overall conformity with the rule of law. However, most (though not all) of that activity blends the gap-filling role with analogs to the work of the other branches, and the priorities that guide reasoning during that work shift accordingly to more closely resemble those of their characteristic branch (concern particularly for some measure of responsiveness or public interest and transparency when that activity is legislative-like, or for proper procedure, predictability, and stability when that activity is judicial-like, though always against the background bias towards substance that belongs to the executive branch itself). Throughout most of this paper, I've used the term rulemaking to refer to the APA's informal notice and comment process, which is mostly an analog of legislative-like activity. The APA, at 5 U.S.C. § 553, provides for formal rulemaking under procedures set out in §§ 556 and 557, in addition to the more common

informal process, but those procedures have fallen out of use since the 1970s when the Supreme Court, in *United States v. Florida East Coast Railway*, 410 U.S. 224 (1973) narrowly interpreted § 553's requirement for when they must be used, and they are much slower and more burdensome than the informal process. Rules made under those provisions are issued by Administrative Law Judges at formal hearings involving evidence and witnesses. Once a decision is issued, it becomes part of the Code of Federal Regulations, in an interesting blend of judicial-like and legislative-like practices. Most agency rules are made in accordance with 5 U.S.C. § 553's requirements for public notice and comment that echo general democratic norms requiring some level of responsiveness and transparency in government(though, to illustrate the point that these norms manifest in a way specific to the branch 553(b)(3)(B) permits an agency to forego notice and comment when it is "impracticable, unnecessary, or contrary to the public interest"), and a requirement that the final rule clearly state its basis (in law) and purpose which echoes government-wide norms of legitimacy and the rule of law.

On the other hand, most informal notice and comment rulemaking activity fits neatly into my category one: activities primarily concerned with making law. Rules made this way under the authority of Congressional delegation are referred to as "legislative rules," and they have the binding force and effect of law. However, there are actually three main sorts of rules that agencies may make as part of the informal process, and statutory mandate is not the only reason an agency may initiate such a rulemaking. Rules may be either legislative, non-legislative (interpretive rules and policy statements), or rules for internal management and procedure. Interpretive rules are rules intended to explain the agency's interpretation of its own rules, and its guiding statutes, to the public—an activity crossing over into category three's mode of interpreting and sharing the judicial branch's attentiveness to predictability in the law, as well as respecting a general norm for transparency. Policy statements give the public notice of how the agency will, for example, use its discretion to prioritize

the enforcement of a particular rule—another measure subjecting characteristically executive action to the same norms for transparency and predictability that figure prominently in expectations for the other branches. An agency may make a rule because Congress has directed that it do so and delegated the necessary authority to it by statute. A rulemaking may also be initiated in response to a petition by the public (§ 553(e)) or (assuming the agency has been granted the discretion to do so) because an agency has identified a problem that requires it to issue a new rule or change an existing one.

The activities an agency may take under the APA themselves run the full spectrum of the possible types of government activities. Those where it acts similarly to another branch show both how it thereby takes the standards for reasoning applicable to those branches into account when doing so (such as by soliciting public opinion when engaged in making law), but also how those standards manifest in ways responsive to the institutional context (the option to forego notice and comment when it is "unnecessary" in the agency's own judgment, for example). The option for an agency to make or change a rule in areas where it has been granted discretion because it has identified some new or unforeseen problem highlights that institutional context and the special attributes executive decision-makers contribute to the business of governing: expertise and leadership. Those rulemakings happen because some official in the agency had the necessary knowledge to identify both the problem and a potential solution, and either that same official or another one assumed responsibility for selecting one when even an expert can't be certain the proposal will have good results when good results are what is most required. They still say little, however, about where those officials—especially the one acting as a leader—should be looking for their reasons when they make those decisions.

One APA rulemaking provision does speak to the matter for a specific class of cases. 5 U.S.C. § 563-566 provides procedures for "negotiated rulemaking" that can be used when an Agency

head determines that a "limited number of identifiable interests…will be significantly affected by the rule" and it is reasonably likely that a representative committee can be assembled to negotiate a rule taking those interests into account to reach a consensus about the proposed rule. These provisions indicate that when possible, an Agency should 1) act in the public interest and 2) seek the input of the affected parties among the public to determine what that interest is. The provision appears on the surface to prioritize responsiveness as a matter of process (a concern that the matter ought to be well-decided), but it is not because the interests—and opinions associated with them—exist that the provision should be invoked. It is because those interests will be "significantly affected." It is because the results matter, and to the greatest extent possible, they ought to be "good"—whatever that means in a particular decision case.

The negotiated rulemaking procedure provides an interesting example of the result-focused nature of executive activity when its officials operate in law-making mode. However, it still doesn't show us an executive activity that takes place solidly in the gap-filling mode, with both institutional and decision-specific bias towards substantively good decisions because of the types of decisions being made and the context they are made in. That means looking elsewhere for examples of how those norms and standards—or others—manifest in the mode of gap-filling and executive leadership or expertise. One source of such examples comes from the military context. Military planning procedures and how the organizations that have designed them describe the needs they address make it clear how the epistemic limits on certainty, results-focused demands for it, and the associated risk burden shape a purely executive leadership reasoning process. The framework for that process is transferable outside of the military context. It suggests norms for those portions of executive branch gap-filling activity that are not practicably susceptible to those imported from other branches.

Every branch of the military has its own version of the operations planning process, which is designed to help leaders and their staffs make plans that reach far into the dynamic and uncertain future—in a context where stakes and risks are high, and results matter to the near[47] exclusion of everything else. (Army Doctrine Publication (ADP) 5-0, 2019; The Operation Process; JP 5-0, Joint Planning, 2017). The Army's version is known as the Military Decision Making Process (MDMP), a creature maligned for cumbersome, tedious rigidity nearly as often as APA rulemaking. The full MDMP consists of a cyclical seven-step procedure from the original receipt of a mission that an operation is to be planned to accomplish, through the execution of that operational plan, and on to the next. Through all steps of the process, the role of the commander (the most important decision-maker involved) is to continuously lead and assess in a progression where knowledge—to the best extent it is available—drives decision. ADP 5-0, the primary Army reference describing the MDMP, depicts the commander's role as a four-segment arrow describing what the commander must do and in what order. (ADP 5-0, figure 1-2, 1-8, 2019, reproduced below). First, she must "understand" the environment that she will be operating in and the problem to be solved. Then, she must "visualize" the desired end state and, in the third step, "describe" it (concretely, in terms of "time, space, purpose, and resources"). Finally, having armed herself with the best knowledge and understanding available before making any decisions, she is to "direct" those under her command.

[47] Near, but only near: the boundaries of international law and constitutional limitations on the use of force still provide at least some element of "proper procedure," as do internal hierarchies and chains of command in most situations.

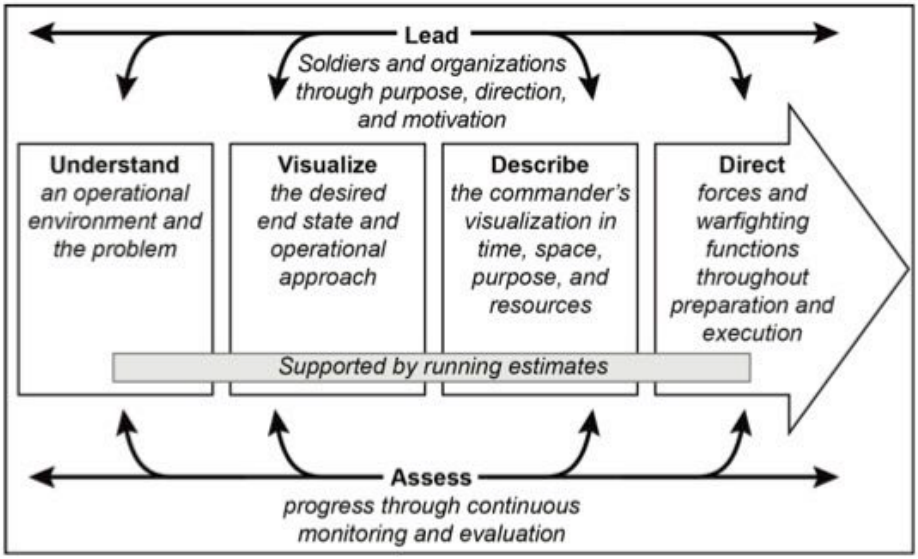

Figure 1, Reproduction of Figure 1-2, ADP 5-0, The commander's role in the operations process.

A detailed description of every step in the MDMP is not necessary here, but its cyclical, integrative approach to the role of knowledge (as well as knowledge gaps and assumptions) in the decision process speaks to the special nature, and therefore special needs, of characteristically executive decision making and its manifestations as both expertise and leadership whether they appear in the context of military decision making, administrative rulemaking, or anywhere else. The MDMP, and the business of military planning and order production in general, takes a systematic approach to gathering, categorizing, and integrating knowledge and awareness of the gaps in that knowledge and their associated risks into the decision process. Information is both the foundation of and incitement for decision. The way this Army decision-making procedure attempts to mitigate risk by establishing a decision process focused on continuously identifying and seeking to close gaps in information is one manifestation of the limited information and unpredictability common to gap-filling activities in any organization. Understanding the role that information and efforts at prediction play in that process, and why they are so important to it, will in turn help make clear what features all such activities share that make them unique gap-filling activities instead of instances of making, interpreting, or applying the law.

The ADP 5-0 diagram reproduced above shows the commander's role in the process as "supported by running estimates." A "running estimate" is more or less what the name would lead one to expect: a continuously updated assessment of a situation, including known facts, any assumptions made, and conclusions and recommendations based upon the combination of those. Running estimates are compilations of what is known, unknown, assumed, and concluded based on the best available epistemic picture. That information feeds the entire MDMP process and the leader's role in it. Since the entire process is conducted under the assumption that there will always be more information to gather, the commander's role, as the leader and key decision-maker involved, includes continuously identifying which pieces of information are most needed. These pieces of information, called CCIRs (Commander's Critical Information Requirements) are associated with "decision points" that denote where (or when) a key decision will have to be made in response to actions taken by some part of the human element involved (enemy, friendly, or civilian) or some other change in the known situation. (ADP 5-0, 2-6).

Even greater nuance is devoted to the effort to systematize the collection and analysis of information in materials promulgated by the military intelligence branch (for example, Army Techniques Publication or ATP 2-33.4, Intelligence Analysis, describes 19 different techniques that intelligence personnel can use to analyze information and assumptions. The running estimate developed by that staff section is the source of the other key example feature that highlights what makes executive decision making special and how it influences the reasoning process. The intelligence staff is responsible for what the Army calls "Intelligence Preparation of the Battlefield" (IPB), a four-step process focused on understanding the environment an operation will take place in and the threats that will be encountered there. In the words of the Army Techniques Publication describing that process, it "assists commanders in reducing uncertainty… Most intelligence

requirements are generated because of IPB and its interrelationship with decision making." (ATP 2-01.3, para. 1-11, p. 1-3).

In an echo of Socrates, the Army (and other organizations that have developed comparable procedures) recognizes that good decision making depends on knowledge, including knowing what is unknown. The IPB process feeds the intelligence running estimate, a crucial part of the command role in planning and the entire planning and operations process. The first step of the process is to "define the operational environment"—a key part of which is identifying existing gaps in knowledge. The fourth step is to "determine threat courses of action" by blending knowledge with analysis, assumption, and risk assessment. The staff's projection of the "most likely" and "most dangerous" of these anticipated threat courses of action[48] are a key element of the planning process and the commander's decision making as the leader of that process.[49] Those projections also become part of the course of action analysis that makes up step four of the MDMP (also called war gaming). In that step, which "helps…identify and mitigate risk…" commanders and their staff "attempt to visualize the flow of an operation, given the friendly force's strengths and dispositions, the enemy's capabilities, and possible COAs, the impact and requirements of civilians in the AO, and other aspects of the situation." (ADP 5-0, 5-36).

Much of the leader decision making in the military planning process revolves around risk that is driven by gaps in information, and efforts to reduce uncertainty and predict threats and outcomes by closing those gaps. The IPB process is just one of many ways that the Army and its sister services seek to reduce risk by reducing uncertainty and making the relationship between the

[48] These could be enemy courses of action but they could also be challenges presented by other types of threats like environmental factors, weather, disease, or unrest in a local civilian population—issues that other staff sections specializing in those areas may also weigh in on.

[49] To include that they are typically included as part of the resulting operations order, and they are an important part of pre-mission briefings and approvals.

discrete areas of the uncertainty that inevitably remains and the risks associated with them into something that can be categorized, analyzed, and reasoned about. Risk and uncertainty, and efforts to mitigate and reason about them systematically, however, are in no way exclusive to either the pre-operation planning process or the military as a whole.

Although it is primarily a tool for leaders and staff who have the time and resources to plan carefully and methodically before an operation has begun, the MDMP and its emphasis—and dependence—on information practices illustrate the challenges leaders face in many contexts, regardless of whether the decision in question is one made at relative leisure early in the military planning process, or an analogous planning process in a civilian agency—say involving environmental risks instead of enemy threats, or the risks of pursuing one policy of criminal enforcement over another. The same elements of epistemic needs and limitation, risk, and uncertainty also appear in many cases where they are only heightened by urgency because an on-the-fly decision that must be made in the moment—whether on the field of battle, at the scene of a crime, or on the site of a natural disaster.

The other common element in this process, shared with other instances of executive leader decision making, is an orientation to future-case considerations (or possible outcomes) that is an inverse analog of the judicial orientation towards past cases. In the MDMP process that I've taken as an example, this is most clear in the practices of "visualizing" the desired end state and iterating various possible courses of action (friendly and enemy/threat) and their likely results as part of the decision process. That practice seeks to know the unknown, make it known to the greatest extent possible, and select from acceptable probable outcomes built around the risks that those uncertainties present. The leader visualizes the desired end state (a set of facts about the state of the world that she wants to bring about) and then considers possible future courses of action and their expected (future) outcomes (possible states or sets of facts her decisions *might* bring about) to select

the one that most closely matches the desired end state. The judge considers the case presented (a set of facts about a past or present state of the world that has led to the controversy she must resolve) and then searches through past cases for sets of facts that may match in a process similar to a commander's consider COA-outcomes that may (to an acceptable degree) match the desired end state. The judge also looks for rules of decision that likewise may apply to the case at hand; a commander's evaluation of potential COAs that could bring an acceptable end state about is, again, similar, though oriented in the opposite temporal direction. If a judge filters past cases to consider whether any match the present case, the commander filters possible future states (cases) to see if any match the currently visualized desired end state—at least to an acceptable degree and within given tolerances for risk and uncertainty. An important difference, though, is the relative importance in each case of the priority placed on the procedure itself and the result of the decision.

In the judicial case, the decision maker operates within an institutional role requiring close continuity with the past and a focus on proper decision-making procedures, even where those preclude the "best" result, for the sake of stability, predictability, and adherence to the rule of law. Only when no past case matches the issue to be decided in the present case may the judge prioritize results over process. Whenever possible, uncertainties and risks should be resolved in favor of continuity with the past.[50] The judge may consider the impact of a decision on either the parties involved or some group of persons that will be affected by the ruling (up to "all citizens" or "everyone" in certain cases). There may be uncertainties about how they will be affected, but those uncertainties are not (usually)[51] the primary concern, and the fact that there are people—a human

[50] With the notable exception of cases where a court that has the power to do so overturns a prior decision judged to be defective—but even then, there must be a connection to established frameworks of law and jurisprudence supporting the decision to overturn.

[51] Unless the judge has concluded that there truly is no law or precedent that applies, and she must therefore fill a gap by using her discretion to, for example, rule on "public policy" grounds.

element—that will be affected and who have brought the dispute before the court—is not itself the main source of the uncertainty the judge grapples with. Those come mainly from the laws, precedents, and facts of the case itself, which were already determined in the past.

In the military case—and executive branch leader-reasoning cases in general, the leader operates in an institutional role oriented above all else towards outcomes. Processes still exist—and as the MDMP illustrates, they can be extremely elaborate—and I do not intend to dismiss their importance. They are important, though, because they are expected to lead to better decisions and, therefore, better outcomes. The uncertainties involved do not come from ambiguities in what legal authorities might control the planning process, whether they apply at all, or what chain of events led up to the need to plan (or decide) in the first place—there is no need to interpret the past (though it should be learned from whenever possible, it is not the *focus* of decision making). The uncertainties and associated risks come from knowledge gaps—which leaders (must work to minimize as much as possible). Of course, the greatest knowledge gaps are those centered around questions about what might or will happen—what outcome will a given friendly COA have? What will the threat do? What will other people in the area do that might change the situation? The future, and especially what other people will do in the future, is the most important uncertainty in play. This is true even if it may be stating the obvious to say that planning and decisions about operations that *will* take place or rules that *will* be made and then enforced are, by nature, future-focused—efforts to choose and bring about desired future states:

> Planning—the art and science of understanding a situation, envisioning a desired future, and laying out effective ways of bringing that future about—is a key activity of the operations process. (ADP 5-0, May 2022, ix).

The MDMP is, by the Army's own description, an effort to cope with the challenges of making high-stakes decisions (to bring that desired future about) in the face of uncertainty from several important sources (including the obvious), but especially the "human element." The Army's

own understanding of the nature of its task describes the adaptive, results-focused, dynamic character that is both the hallmark and purpose of leader reasoning, which fills gaps in the flow of government authority that it would not be possible to address with any static law or rule:

> To understand the fundamentals of planning, Soldiers first must appreciate the general nature of operations. Military operations are human endeavors—a contest of wills characterized by violence and continuous adaptation among all participants.

This human element, and uncertainties that are similar in nature, if different in source, are present throughout executive decision making, whether in the military or outside of it. The human element and results-focused decision procedures designed to reduce uncertainties and their associated risks are likewise pervasive. The Army has the MDMP, other branches have their own processes, and the DoD as a whole has a Joint Planning Process. The FDA has detailed procedures for food and drug evaluation and approval. The CDC has procedures for evaluating the spread of and threats presented by infectious diseases. The NHST has procedures for testing vehicle and roadway safety. In each, as much uncertainty as possible about outcomes is resolved, but the resulting decision must always be made while some unpredictability and risk remain. That decision, however, is not static, especially when it is a case-by-case gap-filling measure (an activity in category four) instead of something more like rulemaking. As additional information, including information about whether the decision was *actually* good (had good results, led to desired outcomes) or not, becomes available, the decision will be continuously evaluated and adapted as necessary.

The key needs and features of the *kinds* of decisions that leaders in the executive branch make when they act as gap-fillers are: they are future-shaping, with risk and uncertainty unavoidable as a result; they may be urgent or in-the-moment, requiring flexibility and limiting time to gather information or consider possible courses of action; they are situation-dependent and too fact-specific that no rule formulated ahead of time or similar past case could address all issues (novelty); likely to have immediate and irreversible impacts (finality); they closely affect and are closely affected

by the human element. Finally, all of these features combined mean that these are decisions that have to be made on the basis of limited information, whether that is because information is not available or there is not time to acquire it. These features of decisions that call for gap-filling by an official acting in a leadership role, where that official may properly rely on their discretion, can be summarized in a list as follows:

- Future shaping
- Risk
- Uncertainty
- Urgency
- Novelty
- Finality
- Human element
- Limited information

These features explain why we need such individual discretion-wielders within a system of laws and institutions in the first place and shape how they reason about those decisions. These features also explain why that reasoning process, though susceptible to the same general democratic and rule of law norms and guided by the same balance points for the use of discretion, will still be results-oriented and focused on what is good, as opposed to well-decided, even in cases where stability and predictability, or transparency and responsiveness, are important values for the decision to be made. Leaders make decisions about what *will* happen, and though they are public *servants*, they are not public representatives or impartial stewards of justice. They make those decisions as themselves, based on their own expertise and judgment (they supply those reasons internally, at least in true gap-filling cases); they are decisive as individuals—their reasons, when reasons must be supplied, are their own. They must decide at which point as much uncertainty and risk have been dispelled as possible before the moment of decision. In doing so, they assume personal responsibility for whatever remains. They often act in and respond to dynamic situations; they must be flexible. Above all, because they make decisions primarily about what will happen in the future in a specific, concrete way, they are outcome-oriented. In response to the unique features of the kinds

of decisions they make, the key features of leader reasoning that make it distinct from representation or judicial reasoning are that it is:

- Future-focused
- Decisive
- Risk assuming
- Dynamic
- Flexible
- Outcome-oriented

To the extent that procedures for such decision making exist, officials are expected to follow them, but adherence to procedure alone will not excuse substantively bad judgment. Likewise, to the extent that information relevant to the decision is available, officials are expected to avail themselves of it—but incomplete information is never an excuse for a failure to act at all, and again epistemic diligence alone will not excuse substantively bad judgment.

This is why, to return to Section III's series of balance points that guide the use (or limitation) of discretionary authority, executive branch category four gap-filling leader decision making by nature prioritizes the side of those dichotomies focused on substantive goodness. The democratic norms of transparency, responsiveness, and solicitude for the public interest that appear as guiding norms for representatives remain relevant. They are prioritized and manifest differently (transparency as to the operations of the CIA, and public accountability for them, for example, are quite different than transparency as to how a representative has voted and its impact on whether he is reelected), but they do not disappear. Likewise, general rule of law norms that require discretion to be used only when it has been properly granted, and no other more established source of government authority already determines the question, continue to apply. However, adhering to them does not require the same sort of focus on stability and corresponding deference to past cases from executive decision-makers that it does in the judicial branch because of the difference in institutional roles. Instead, leaders in the executive branch must engage in decision making that is

both accurate, agile, dynamic,[52] and carried out with respect for the inherent risks of such decisions. At the start of this paper, I noted that we ought to be concerned with how the authority these leaders wield fits into a larger framework of representative democracy since they often aren't directly answerable to the citizenry and their activities may even be secret. To some extent, there is an uneasy tension here that cannot be satisfactorily resolved other than by reiterating the fact that such decision-makers are necessary if the government is to carry out all of its purposes—especially those involving national security—adequately, and that by these standards their discretion is to be employed *only* when necessary. At least to an extent, however, it is also true that the tradeoff of being entrusted with the discretion to wield government authority as an individual decision-maker in these kinds of cases is that the leader making one of these decisions might be held personally responsible for the results. Sometimes that responsibility could take the form of professional sanction. Sometimes it might see an official explaining him or herself to a board of superiors or even to Congress. Sometimes it might take the form of public censure, and sometimes—assuming the official in question is appropriately motivated and conscientious—it might be a matter of self-recrimination.

There are times when the decisions made by the legislature or judiciary will be ones where procedure, that the decision be well-decided whether that means that it be responsive to popular opinion or consistent with past precedent may arguably be more important than that the be substantively good. In the executive branch, when it operates under a grant of discretion in its gap-filling role this is never the case. A decision made in that capacity according to all the best procedures that nonetheless had disastrous results would be undesirable in the same way that a medical prescription given on the basis of the best, most recent scientific evidence but which caused

[52] Again, in contract to judicial decisions which are expected to grant parties and the public "repose."

an unforeseeable and fatal allergic reaction would be undesirable. Likewise, a decision arrived at according to truly questionable or even objectionable processes, but which had the best of all possible results would be preferable—even if not entirely desirable in its own way—in much the way that the accidental dispensation, through negligence, of the wrong prescription which nevertheless miraculously effected a cure would be. These cases are not unlike the high-stakes decisions leaders make in the face of unpredictability, and our standards for evaluating such decisions must take those features into account. While we may rightly place blame on the party responsible for a disastrous result caused by negligence, there is no procedure we could devise that would avoid that same outcome had it been the result of chance and we would not reasonably suggest that the prescribing physician should have waited until all possible data that could have rendered it predictable were available before rendering any treatment at all. It is impossible to ever know everything, and so it is impossible to eliminate every risk—but we may tend to feel culpability or wish to place blame nonetheless when procedures were proper but the result was still bad. It's not that procedure never matters in such cases, but it is never what matters the most.

These features, especially personal investment in the burden of risk and uncertainty, make the reasoners involved not just officials, agents, decision-makers, or even "commanders," but leaders per se. The institutional roles and priorities of officials in the legislative branch dictate that they are first and foremost to act as representatives, whatever that means in a given context. They may prioritize proper procedure and substantive goodness, and base their law-making decisions on either external or internal reasons differently at different times, so long as they act according to some concept of the public interest and adhere to the norms appropriate to whatever method of representation they are engaged in. The institutional roles and priorities of judicial officials require that they often give priority to decisions that are procedurally well-decided because they serve an important stabilizing function in government that requires predictability about what the laws and

legal processes are. Where it is the role of legislative officials to represent the people while making laws, and of the judiciary to maintain the impartiality and stability of the laws that they interpret, the role of officials in the executive branch likewise requires a bias towards decisions that are substantively good when they act as leaders who must fill gaps left in the law by the other branches or react to the inherent unpredictability of changing circumstances and the changing needs of a living population over time. Their purpose is not to create the legal framework of an enduring democratic system of government, as is the legislature's, nor is it to maintain the integrity and stability of one, as is the judiciary's. Their purpose is to respond to specific needs of those institutions, and the people they serve, that could not be foreseen to be codified and that cannot wait on procedure.

Works Cited

Alexander, L. & Sherwin, E. (2008). *Demystifying Legal Reasoning.* Cambridge University Press.

Bonfield, A.E. (1972). Military and Foreign Affairs Function Rule-Making under the APA. *Michigan Law Review, 71*(2), 221-358.

D'Amato, A. (2011). On the Connection Between Law and Justice. *Northwestern University School of Law Faculty Working Papers,* http://scholarlycommons.law.northwestern.edu/facultyworkingpapers/2.

Dworkin, R. (1963). Judicial Discretion. *The Journal of Philosophy, 60*(21), 624-638.

— (1985). Law's Ambitions for Itself. *Virginia Law Review, 71*(2), 173-187.

— (1986). *Law's Empire.* Harvard University Press.

— (1967). The Model of Rules. *The University of Chicago Law Review, 35*(1), 14-46.

Epstein, J. (2018). Drugs that Shaped the FDA: From Elixir Sulfanilamide to Thalidomide. *Bulletin of Historical Chemistry, 43*(2), 102-110).

Folk, T. (1986). The Administrative Procedure Act and the Military Departments. *Journal of the National Association of Administrative Law Judiciary, 6*(3), 109-143.

Hart, H.L.A. (2012, originally published 1961). *The Concept of Law*, Third Edition. Oxford University Press.

— (1958). Positivism and the Separation of Law and Morals. *Harvard Law Review, 71*(4), 593-629.

Jacobs, S. (2009). Crises, Congress, and Cognitive Biases: A Critical Examination of Food and Drug Legislation in the United States. *Food and Drug Law Journal, 64*, 599-629.

Raz, J. (1979). *The Authority of Law: Essays on Law and Morality.* Oxford University Press.

—(1972). Legal Principles and the Limits of Law. *The Yale Law Journal, 81*(5), 823-854.

—(1986). *The Morality of Freedom.* Oxford University Press.

Schauer, F. (1987). The Jurisprudence of Reasons. *Michigan Law Review, 85*(5), 847-870.

—(1991). *Playing by the Rules: A Philosophical Examination of Rule-Based Decision making in Law and in Life.* Oxford University Press.

—(2009). *Thinking Like a Lawyer*. Harvard University Press.

—(2008). Why Precedent in Law (and Elsewhere) Is Not Totally (or Even Substantially) about Analogy. *Perspectives on Psychological Science, 3*(6), 454-460.

Shapiro, S. (2007). The "Hart-Dworkin" Debate: A Short Guide for the Perplexed. *University of Michigan Law School Public Law and Legal Theory Working Paper Series, Working Paper 77*.

Stevens, K. (2018). Reasoning by Precedent—Between Rules and Analogies. *Legal Theory, 24*(3), 216-254.

Sunstein, C. (1993). On Analogical Reasoning. *Harvard Law Review, 106*(3), 741-791.

Government Publications

Headquarters, Department of the Army. Army Doctrine Publication, (July 2019). (ADP) 5-0: The Operations Process. (Distribution unlimited/publicly available at https://armypubs.army.mil/ProductMaps/PubForm/Details.aspx?PUB_ID=1007409).

—(January 2020). Army Techniques Publication (ATP) 2-33.4: Intelligence Analysis. (Distribution unlimited/publicly available at https://armypubs.army.mil/ProductMaps/PubForm/Details.aspx?PUB_ID=1008410).

—(March 2019). ATP 2-01.3: Intelligence Preparation of the Battlefield. (Distribution unlimited/publicly available at: https://armypubs.army.mil/ProductMaps/PubForm/Details.aspx?PUB_ID=1006342).

United States Space Force. (December 2021). Space Doctrine Publication (SDP) 5-0: Planning for Space Forces. (Publicly available at https://media.defense.gov/2022/Jan/19/2002924107/-1/-1/0/SDP%205-0,%20PLANNING%20(20%20DEC%202021).PDF).

United States Senate. Rules of the Senate. (Available at https://www.rules.senate.gov/rules-of-the-senate).

Congressional Research Service. (2008). House and Senate Rules of Procedure: A Comparison. (CRS Report RL 30945, available at https://crsreports.congress.gov).

Paper 3: A Different Kind of Responsibility Gap: Trust and the Burden of Risk as a Limit on Military Automation

Public service is a public trust—5 CFR § 2635.101 –Basic obligation of public service.

…reposing special trust and confidence in the patriotism, valor, fidelity, and abilities…—US Army Officer Commissioning Certificate

This paper argues that the most far-reaching obstacle to the widespread implementation of autonomous weapons systems is a problem that's not so much about autonomous weapons systems themselves as it is about the automation of government decision making. In asking when and how we may permissibly employ automated weapons of any kind, we are asking when and how—and which—decisions may be delegated to computers instead of human beings. We need to think carefully about when, where, and why we should or should not replace delegation of decision-making authority with automation. One important consideration that we should consider is how automation may affect our ability to trust in military decisions and decision-makers in the necessary ways. Even if we can implement automations that reliably behave in the correct ways, this is not necessarily the same as trusting them the way that we trust human beings making the same sorts of decisions. The question is whether anything important is thereby lost. I will argue that the answer is yes. The need to maintain that trust in the military process is one important guidepost for how we should and should not incorporate automation as technology advances. It may be the case that an automated system could more accurately and efficiently perform the functions we currently trust to human military servicemembers, and not only at the lowest levels of combat. This is often cited as a benefit offered by autonomous weapons, along with the argument that offloading combat decision making this way will alleviate the heavy psychological burdens currently incurred by human combatants. However, this appealing benefit highlights what is problematically lost when we automate government decisions. I do not mean to suggest that the negative impact of combat on

mental health is itself in any way desirable, but there is also a part of that "psychological burden" that is a necessary byproduct of a human agent's properly grasping the important responsibilities that come along with their authorization to engage in state-sanctioned violence. The burden of experienced responsibility inherent in their treating the decisions made under that authority with the seriousness they deserve is something that automation cannot replicate.

Section one of this paper summarizes the concerns raised in opposition to and in favor of autonomous weapons. Section two proposes revisions to the operative distinctions so far drawn in that debate. Section three argues that the discussion should be reoriented to address a wider problem with delegating military decision authority to computerized systems of any kind—including not only automated weapons but also all phases of strategic and planning process since the arguments that computers can make "better" decisions about targeting and engagement apply just as much to decisions about gathering and analyzing information, selecting one strategy over another, creating plans to carry those strategies out, etc. Section four discusses the problem of trust as it applies to the question of automation, arguing that there are affective and normative elements—psychological elements—to trust and trustworthiness that are necessary to the military decision process and that they cannot be properly bestowed on AI. Finally, section five explains why human leaders are required to maintain this trust even if they are more fallible than automated systems, and their replacement could alleviate certain psychological burdens of conflict.

Section I. Automated Weapons, or Automated Decisions?

This section will briefly review existing arguments favoring or opposing the use of Autonomous Weapons Systems (AWS)[53] or Lethal Autonomous Weapons Systems (LAWS). That

[53] The discussion will often refer to AWS generally, as opposed to only Lethal Autonomous Weapons Systems (LAWS). While there are some issues addressed by the literature I will engage with that pertain specifically to lethality, I have chosen not to so limit this discussion because the larger question the paper addresses has to do with the delegation of either the state's authority to employ violent force whether the impact is lethal or not, or state decision-making authority in general.

discussion forms the context in which I will argue for reframing the central question as one of delegation. There are good reasons to believe that we make an important conceptual mistake by focusing that discussion too much on autonomous weapons systems as weapons systems or physical systems per se (Crootof 2018). Focusing exclusively on whether we can or should create robotic or otherwise automated offensive weapons systems capable of complying with the legal and moral constraints on modern warfare tends to obscure important underlying issues surrounding the automation of military and government decision making. Questions about whether killer robots or computerized assassinations can ever be morally, legally, and politically justified are just some specific manifestations of more generally applicable questions about automated decision making and the appropriate flow of government authority. The principles that will answer them apply regardless of whether the locus of automation in question is a semi- or fully autonomous system as a matter of design, a lethal or non-lethal system, put to offensive or defensive use, or a weapons system or any sort of "embodied" system at all. Even assuming all the commonly discussed concerns about whether we can, permissibly or practically, create "fully" autonomous weapons are resolved in favor of automation, these other issues remain. They merit serious discussion because they set the framework within which military and government automation should be considered.

Arguments against autonomous weapons systems may be either practical or philosophical. Practical objections include such claims as that we lack the capability to make one that successfully complies with legal and moral principles for the conduct of warfare. Philosophical objections include such claims as that, even if they do function properly, they violate some intangible normative requirement for justified application of force, that they will make violent conflict more likely because they reduce the associated political costs in an undesirable way, or that they introduce insurmountable gaps in our ability to assign legal or moral responsibility for mistakes or other failures to comply. Determinations as to which weapons systems should be ruled out by these

objections tend to designate the decisive characteristic as one of a handful of distinctions. Those frequently raised are:

- Whether a system is semi- or fully autonomous (or whether there is a "human in the loop" or "meaningful human control"); or
- Whether a system is a *lethal* autonomous weapon system (LAWS) or just some other variety of autonomous weapon system (AWS).

However, the substance of the associated objections against those systems deemed unacceptable according to these distinctions appears in practice to have less to do with any immutable characteristic of a system itself and more to do with the purpose they are used for: Whether they are put to defensive or offensive use.

The prevailing distinctions between *systems* exhibiting differing degrees of autonomy, or those designated lethal versus non-lethal, fall short of addressing the substantive concern at work. The concern is which instances of our delegating decision making away from human agents are acceptable and which are not, not which types of automated systems are acceptable and which are not, or what degree of such delegation is acceptable for a given type of decision-making authority. The analysis, I believe, is not unlike the designation by law and policy of "inherently governmental functions"[54] that must be performed by federal personnel rather than contractors. There are likewise some "inherently human functions" that must be performed by human agents, rather than the most sophisticated, accurate, and efficient non-human systems. To properly address that larger question about whether and when we ought to accept automation in place of delegation to a human agent, however, requires a grounding in the existing weapons-system-focused debate.

For all practical purposes, the discussion about military use of AI and autonomous weapons has moved far beyond "if" or "may" to "how," "when," and "why." Computer-assisted analysis and

[54] That distinction is addressed, for example, in subpart 7.5 of the Federal Acquisition Regulation (FAR).

limited uses of automation are already in use in various ways. C-RAM (Counter Rocket, Artillery, and Morter) systems and the U.S. Navy's Phalanx CIWS (Close In Weapons System) have been in defensive use for over a decade (Scharre and Horowitz 2015).[55] Because a human operator must provide an initial set of targeting criteria, some definitions would class these systems as only semi-autonomous—but once programmed and activated, they detect and engage threats that meet the targeting criteria without further human guidance. However, they are designed and used exclusively for defensive purposes (to defend land bases or ships). Steps in the development of effective offensive autonomous systems have garnered much more attention. They are the subject of enough alarm that, in addition to academic discussions about the ethics and legality of their use and design, there has been an ongoing international conversation about whether they should be preemptively banned (Horowitz 2016, Sayler 2023).[56] Objections against them include that it will not be possible to design systems that adequately comply with the Law of Armed Conflict either as a matter of use or design. Concerns centered around their use generally concern their ability to distinguish adequately between acceptable and unacceptable targets. Practical concerns around their design include that technology is not yet advanced enough to produce a usable system or that usable systems may still be too vulnerable to hacking or hijack. Other design-focused objections cite the Martens clause, which requires that in the absence of any clearer provision of statute or treaty, the conduct of warfare should be constrained by "principles of humanity and the dictates of conscience" (Sayler February 2023).[57] The issues to be overcome, then, can be roughly divided into two. One set

[55] These systems work by detecting incoming rounds according to targeting criteria set by human operators. Once those criteria are set, they can autonomously neutralize threats—meaning that if a target meeting the set criteria is detected, the system will autonomously fire rounds of its own to destroy the incoming projectile. Similar systems are used by many nations. *See, e.g.,* https://iamd-coe.org/focus-areas/counter-rockets-artillery-and-mortars-cram/ accessed 29 June 2023.

[56] https://news.un.org/en/story/2019/03/1035381 accessed 29 June 2023.

[57] The Martens Clause appears in the preamble to the 1899 Hague Convention. Customary international law is the idea that principles and customs accepted by most nations making up the international community are binding as part of international law. Appeal to the Martens clause relies on two presumptions: first, that we are correct to found legal and political objections to AWS on moral ones, and second that a majority of the public conscience, as well as principles of

of issues includes more straightforward questions about, first, the possibility of the design and widespread use of AWS and, second, with the permissibility of their design and use. The first set has to do with compliance: whether we can or should create and use *fully* autonomous weapons systems that meaningfully comply with legal and ethical standards for the conduct of armed conflict. They are genuine questions *about* the law and ethics of warfare. In contrast, most of the latter set of issues are questions about weapons design, weapons use, or any other principle of the Law of Armed Conflict or the ethics of warfare only on the surface. At a deeper level, they are questions about military *decision making* and automation's proper place within it (if any). At yet a deeper level still, they are questions about automation as a novel form of delegation and whether and where limits exist on this type of delegation of state authority.

Though those citing the Martens clause in the context of autonomous weapons are most often using it to express concerns that those weapons could violate the clause, many who favor their use do so because they believe that AWS will be better at complying with the legal and ethical constraints law, custom, and morality impose on warfare than fallible human agents are (Arkin 2008, 2010, 2012). If this were the case, the argument goes, they are not only morally permissible, they are desirable, and we might even have an *obligation* to employ them. This position might even seem to track with a claim that the Martens clause, and the principles it invokes, similarly permits, or even requires, the use of AWS if it is true that they will be better at deciding when to use force than human agents.

For the sake of argument, let us suppose all the technological, legal, and philosophical difficulties having to do with function and design can be resolved so that we are able to create a well-functioning robotic system equipped with artificial intelligence that has been successfully

humanity, are opposed to their use—a viewpoint that is by no means settled even if there were agreement that the clause, which is an extremely broad pronouncement in the preamble rather than a substantive treaty provision, is in some way enforceable.

equipped with all necessary legal and moral principles and the ability to apply them in all possible cases. Given a system that would then arguably be an objectively superior military reasoner than a human servicemember, we must address the argument that we are not only permitted, but *obligated*, to replace human soldiers and commanders with such systems. Because they theoretically will not err, or at least will err less than human agents do, and because they cannot experience physical *or* psychological injury, doing so will both improve the overall legal and moral correctness of decisions made in war and reduce suffering on both sides of a conflict.

What possible reasons could we give to claim that less-accurate decision making is desirable or even permissible, in that case—or that it is likewise desirable, or even permissible, to continue to allow individual servicemembers and leaders to bear the psychological burden of even correct decisions to employ destructive force during the conflict? For one, appeal to the objective superiority of the "decisions" made by automated systems does not overcome objections that derogation of human responsibility for such morally significant decisions as those made by AWS undermines human dignity and fundamental principles of moral accountability (Asaro, 2013, 2020). This objection alone, though, does not address issues of institutional normative health or leader decision making touching on trust and trustworthiness that must be part of any discussion involving not only morally significant decisions, but significant *government* decisions.[58]

Asaro's account of such objections turns in part on adoption of the ICRC (International Committee of the Red Cross) working definition of an autonomous system, which is one capable of selecting and engaging a target without "meaningful human control," or "human intervention."[59]

[58] There is popular international support for a pre-emptive ban on systems lacking meaningful human control, though the specifics of when a particular degree of human supervision is or is not "meaningful" is not well settled in any of the academic, legal, or international communities. (See, e.g., Schmitt and Thurner, 2013, arguing that there is no need to preemptively ban AWS because we should reasonably expect to see only "limited autonomy" systems in the near future, and those can comply with existing law).

[59] https://www.icrc.org/en/document/what-you-need-know-about-autonomous-weapons accessed 29 June 2023.

This is often contrasted with what are called "human in the loop" systems, where at least at some level, a human being is involved in supervising the machine—such as a patriot missile requiring an operator to press a button confirming a target, or even the CIWS/C-RAM requiring human-generated targeting criteria. Where the line is drawn depends on what sort of human control one considers "meaningful" (*See, e.g.,* Cavalcante Siebert et al. 2023, Horowitz 2016a, Horowitz 2016b).

Additionally, the more successful in creating a system with sophisticated and robust autonomy we are, the more relevant a set of worries about more abstract challenges facing any artificial intelligence or autonomous system (weapon or otherwise) becomes.[60] Those problems center around whether we can or should trust autonomous systems, and whether they will leave problematic moral and legal "responsibility gaps" in their wake (Oimann 2023, Santoni de Sio and Mecacci 2021). Limits or barriers to trust include opacity in both the learning and decision processes and challenges in assigning legal or moral responsibility, or both, for the actions of these systems (Kaur et al 2022, Ryan 2020, Burrell 2016, Oimann 2023). When these questions are focused on the issue of AWS, specifically, they ask whether we can or should entrust the authority to engage in state-sanctioned violence at the lowest levels to autonomous weapons systems at the lowest levels of such decision making (those of the immediate identification and engagement of individual targets by "robot weapons")—but why, assuming their capabilities really are objectively superior to those of human beings—would we stop at the lowest levels? The military leader issuing orders from a command post is not so much less likely to be prone to error or emotional influence than the soldier a robot might replace on the ground that we shouldn't also think an automated command system

[60] A separate set of objections against AWS have to do with concerns that they will in fact *not* be reliably more accurate, or better on some objective metric, than human reasoners regardless because of practical difficulties in creating well-functioning robotics systems to be the interface between artificial intelligence and the physical world that leave potential systems more resembling a Roomba than the Terminator, and the legal and philosophical difficulties in creating *any* computer system that can reliably interpret and apply the entirety of a complex legal or moral code (Prakken, 2017, Anderson and Anderson, 2015).

could be more accurate than an automated weapons system, since we've assumed all of the technological and conceptual barriers to creating a well-functioning and sophisticated fully autonomous system have been overcome.

Yet, here the remaining problem becomes clear: the discretionary authority granted to military leaders is accompanied by the special responsibility these leaders assume for the risks associated with their high-stakes decision making. The psychological burden that accompanies that responsibility is an important constraint on their power. The fact that they are subject to such a burden—that they handle that power with an attitude appropriate to its nature—is necessary for the health of the trust relationships that they must maintain, both internal to the military institution (among those who must work together to make such a decision) and external to the institution (in relation to institutional channels through which delegated authority flows, and between military and government institutions and the public).

Section II. When We Talk About Using Autonomous Weapons, Which Decisions Are We Talking About Delegating?

I've argued above that we ought to reframe the discussion of whether autonomous weapons are permissible in terms of whether it is appropriate to entrust certain types of government authority to automated systems. As part of that argument, I noted that there are differences of opinion as to whether semi- ("human-in-the-loop") or fully autonomous systems are different in relevant ways,[61]

[61] There is no accepted standard for what constitutes autonomy, nor for what constitutes the meaningful human control that many argue renders an automated system less-than-fully autonomous (and, according to many, morally or legally defensible). The International Committee of the Red Cross (ICRC) currently applies the term to a weapon system that can "select and apply force to targets without human intervention." The ICRC includes systems that make these selections on the basis of a general "target profile" provided by a human being prior to the system's activation. The focus of its concern is that the system can select individual targets independently: "the user does not choose, or even know, the specific target(s) and the precise timing and/or location of the resulting application(s) of force."[61] Since knowledge as to the location at which force is to be used seems to matter for the ICRC's criteria as well, it's unclear if defense systems like the Phalanx and C-RAM that are given only target *profiles* would still fall short of that definition (which would otherwise seem to include them, but in context appears to contemplate mobile, offensive systems) or not since there can be no mystery as to the location at which force has been applied, and it is likely the target engaged will be known.

Current US policy, as reflected in DOD Directive 3000.09, "Autonomy in Weapon Systems," (https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf), describes a lethal autonomous

and likewise as to whether we ought to treat those systems that are lethal differently than those that are not, or that there is a relevant difference to be found in whether a system is used defensively or offensively. Those distinctions are often taken as operative in determining which proposed systems are acceptable and which are not. These distinctions, however, are not only *not* dispositive of the legal, moral, and political issues at hand: they are distractions and confounders. They continue to dominate the narrative, however, because they each touch on some part of the underlying factor that is decisive. They are different ways of asking about the degree and importance of the decision making that has been automated in each sort of case. An automated system capable of destroying incoming projectiles raises fewer objections not because it is less autonomous than an autonomous drone sent to attack the hostile person launching those projectiles, and not because it is objectively less lethal (since these systems are often equipped with missiles). It is less of a concern because the right to act in defense is much more freely delegated to even the lowest levels of authority than the right to select targets or even to attack a pre-selected target. The operative difference is whether a system is able or allowed to do things that would, were it a human agent instead, require independent judgment, discretion, and responsibility for risk—and if so, how much.

---

weapon as one that can select and engage a target without human intervention, as opposed to one in which a human operator can intervene before engagement or stop engagement.[61] It defines semi-autonomous systems as those which autonomously engage targets or groups of targets that have been selected by a human operator (such as the Phalanx or C-RAM). It requires, however, that all automated systems (fully or semi-autonomous, lethal or not) be designed to allow for an "appropriate" degree of human oversight and intervention in the use of force (Sayler May 2023). This is essentially an alternative phrasing of the requirement for meaningful human control, understood as the opportunity to stop engagement when such intervention is deemed necessary. Interpretations of what is appropriate are likely to vary so widely with context and opinion as to render the restriction hopelessly subjective. Further, a definition of an "autonomous" weapon as one that engages a target without intervention, in contrast to one where an operator is able but not required) to intervene implies that there is no opportunity or ability for intervention in the first case. This raises the question of how the first sort of system could reasonably comply with the later requirement that there be "appropriate" oversight and intervention unless there is theoretically some situation in which the "appropriate" degree is "none."

Looked at this way, the attempt to distinguish between fully and semi-autonomous weapons and yet require oversight and intervention for the use of both seems to indicate that while the policy may permit the use of systems *capable* of full autonomy, it does not necessarily permit their fully autonomous *use*. That's not a distinction about whether we can make or use robots capable of killing people or destroying property, it's about whether we can trust the authority to do so to a system lacking human agency. These are weighty official decisions; can we pass them off to something that doesn't truly "decide"?

If the relevant question is whether we ought to delegate authority to computers to "decide" to employ destructive force, we should include in the category of fully autonomous systems any system capable of identifying and engaging a target without *in-the-moment* human input, regardless of the context. The question of whether human control is meaningful or not is whether there is human control over each instance of force. This is a more expansive definition than some, including the current US Policy, in that it does away with the distinction between semi- and fully autonomous, or those (for example, Roff 2014) requiring the ability to choose targets independently based on learning and adaptation. It would consider a system that proposes a target but requires human confirmation to be non-autonomous. It would consider systems like the Phalanx and the C-RAM to be autonomous because such a system can identify a proper target for the use of force and employ that force without human input as to a particular engagement. I have chosen to use the word in this broader sense because the foundational consideration about the flow of decision-making authority applies to either sort of system, and the reasons one kind is deemed acceptable and another unacceptable are likewise about that flow primarily and only about the design or capabilities of a given system incidentally. The distinction between lethal and non-lethal systems is another distinction often assigned moral relevance. Taddeo and Blanchard, for example, argue that a different (greater) degree of moral risk-taking is acceptable for non-lethal systems but not at the greater level of force they attribute to lethal systems (2022).

These two distinctions are more distracting than helpful for understanding military automation as a question about decision-making authority because they tend to encourage thinking of the issue as being about the design or capabilities of a given system. The third distinction, between purely defensive systems and those capable of offensive applications, begins to clarify matters. Much like the distinction between semi- and fully autonomous weapons is often cited as the relevant one, those who call for bans on automated military systems tend to refer to Lethal

Autonomous Weapons Systems (LAWS) instead of AWS, implying that lethality, as opposed to weaponization, is the most important issue such that the morally and legally relevant distinction has to do with the degree or effectiveness of violent force applied. They are not distinguished as "human targeting" versus "structure targeting" weapons systems, after all. Nonetheless, the objections given to LAWS as worse or more concerning than AWS sound less like arguments against the *potency* or *efficacy* of the force used and more like arguments against its *purpose* (offense, as opposed to defense). Likewise, those systems designated as concerning because they are fully, as opposed to semi-, autonomous also happen to be those systems that could be put to offensive use. Alternatively, it may be that it is offensive use that requires the greater degree of autonomy. A defensive system deployed with no human oversight at all might raise objections, much like tort law's infamous spring-gun[62], if we believed it incapable of properly identifying threats and responding to them with the appropriate degree of force but those objections would be the sort of design-and-function problems that have been set aside for this discussion. Assuming that it *was* so capable, such a system does not raise the same important questions about decision making that an offensive system with the same level of autonomy and capacity for destruction would. Instead, it is the decision to *act* offensively instead of to *react* defensively that makes the difference.

That difference is how much and what kind of authority has been transferred from a human decision-maker to the system in question. It is still a question of which decisions can properly be automated and how far that automation can reach. A base or ship defense system is "non-lethal" to the extent that its targeting criteria are programmed to include only incoming projectiles, but that does not make it any less an autonomous computerized missile system—potentially a highly deadly

[62] The so-called "Spring-Gun Case," Katko v. Briney, 183 N.W. 2d 657 (Iowa 1971), established the precedent that a home-owner can be held liable for injuries to a trespasser that result from a booby-trap on their property. The Brineys were homeowners who had set up a spring-gun—a shotgun rigged so that it would fire when the door was open—in an abandoned house located on their property to deter intruders. Katko was a trespasser and would-be thief was inured when the trap performed as intended. Katko successfully sued the Brineys for his injuries.

weapon. Likewise, there does not seem to be much point in discussing the relative acceptability of an autonomous UAV armed with rubber bullets or an automated LRAD (Long Range Acoustic Device) for the sake of describing an AWS that is not a LAWS. The LRAD emits a high-power frequency that can be painful and may be powerful enough to cause permanent hearing loss. Rubber bullets, though generally considered "non-lethal," are nevertheless also capable of causing permanent injury, and they are, in fact, lethal in approximately 3% of cases involving their use for crowd control (Haar RJ, Iacopino V, Ranadive N, et al. 2017).

Any weapon, if it is effective, is potentially lethal, whether by design and intent of use or not. Even the LRAD may cause a dangerous stampede as people attempt to escape its effective range in which persons could be trampled to death, just as much as bombing an unoccupied enemy building could cause calling debris to crush a passing civilian whose presence no one could have predicted. The doctrine of double effect does not alter the outcome for the deceased. Any offensive action is a potentially destructive action, or there would be no point in our undertaking it. Any potentially destructive action is potentially lethal action—and it is unlikely we will construct specifically non-lethal offensive systems—and it turns out the most pressing questions deal not with lethal *potential* but with *offensive* use.

Michael Horowitz's research into public opinion on the debate over so-called lethal automated weapons provides empirical support for the claim that the public at large is less opposed to the use of automation for defensive purposes and more for offensive purposes—and when their use is presented as a way to reduce danger to human servicemembers as opposed to an improvement in our offensive capability (Horowitz 2016b).

A wall deflects projectiles without human intervention as well, but we would certainly not call it semi-autonomous. Even the lowest-ranking, least experienced military member is not denied her right to employ force in defense of her own life without further authorization, though her

discretion to select and engage targets offensively may be limited by the Law of Armed Conflict, Rules of Engagement, and command authority down to the lowest levels.[63] She is capable of fully autonomous offensive lethal action, but she is not authorized in many situations to take it. The legal and moral defensibility of her actions in war, and their political and social acceptability, don't have anything to do with her capabilities; they have to do with her permissions and her purpose.

Questions about how these decisions ought to be made are not questions about lethality versus non-lethality. Whether the use of AWS to target a particular individual raises two separate questions, one about the acceptability of automating the decision to employ destructive force of any kind, and one about the acceptability of targeting that particular person, method immaterial and assuming authority for the use of force in general has been properly delegated—as opposed to one question about whether it is acceptable to automate the decision to employ destructive force against a particular person. The authority to employ force without legal or moral sanction is central to the authority of the political state. That authority is delegated to commanders and servicemembers with varying degrees of discretion and restriction within the framework of the Law of Armed Conflict, Rules of Engagement, and strategic and operational plans and orders. Historically, the furthest endpoints of that delegation have been the individual servicemembers who must finalize the plans, issue the orders, or pull the triggers on the field of battle.

In this section, I have argued that we should focus on questions about what authority, and how much, may be diverted from human agents as the most pressing issue surrounding the use of AWS as opposed to how autonomous (or not), or how lethal (or not) they might be. This is because the introduction of AWS introduces new endpoints for the flow of that authority, but so, potentially,

[63] There is disagreement about whether the right to personal self-defense is incorporated into the Law of Armed Conflict or instead rooted in common principles of domestic law allowing that right. For an explanation of why the issue is unclear, though there is no dispute that such a right exists, see Cherry and Rizzotti, "Understanding Self-Defense and the Law of Armed Conflict," https://lieber.westpoint.edu/understanding-self-defense-law-armed-conflict/ accessed 5 July 2023.

does the introduction of technology that could be used for automated intelligence gathering and analysis, automated target proposal (as opposed to selection), more extensive automated collateral damage analysis, and the automated evaluation, or even selection, of strategic options. Some of these capabilities are already in use. The modern command post is an assembly of screens and computers. The targeting process employs software that models collateral damage estimates. The planning process likewise incorporates computerized modeling to assess the impact of proposed or expected courses of action. The leap to also automating selection or approval at certain steps in planning and command processes as technology improves is a much smaller one than we might imagine if we think of military automation only in terms of robot soldiers and the like. In the next section, I'll address these possibilities as another reason to think focus on any particular version of autonomous weapons alone causes us to ignore a larger question.

Section III. Why Settle for a Terminator When You Could Have Skynet on Your Side Instead?

I've said above that we should think of the problems AWS present for legal, moral, and political norms as primarily questions not about a new sort of weapon, but instead about a new sort of delegation. One advantage of this shift is that it allows us to reconceptualize the question of military applications for artificial intelligence as more than a special legal and ethical problem having to do with the use of violent force by robotic autonomous weapons systems, although this is the prevailing view problem since existing systems such as the Phalanx and C-RAM that don't fit our image of "robotic" per se are still applications of autonomous computing power linked to specific purpose-built physical weapons systems, as are loitering systems like the Turkish Kargu-2, which is

capable of both remote control and autonomous operation[64] or the Israeli Harpy drone,[65] which targets radar emissions in a pre-programmed area.

I stated above that current and foreseeable technological capability for robotic autonomous weapons provides systems that remain more like a combat Roomba than a Terminator. It is true there are likely to be advances in our ability to successfully integrate computing power with more advanced sensing and actuating technology. Even Roombas have improved since their introduction. However, to continue to abuse the analogy to 1990s popular culture, the more likely effective, efficient, and concerning applications of machine learning, algorithmic analysis, and whatever other automated computing capabilities we wish to describe as "AI" in the military context would be better described as Skynet. The aspect of AWS that is both most interesting and most challenging is entirely non-physical. It's not the machines themselves that raise concerns, it's their programming and computational power.[66] Rebecca Crootof (2016) points out that focusing only on physical weapons ignores cyberweapons as a class of semi- or potentially fully autonomous weapons with effects ranging from the disruptive to the destructive or even deadly.[67]

Much like focusing only on embodied weapons ignores important questions about cyber weapons, focusing exclusively on weapons at all ignores many other potential and likely points of automation for observation, monitoring, modeling, analysis, and decision making. We talk about replacing soldiers with robots, but what about commanders and planners? No weapon is launched

---

[64] See "Letter dated 8 March 2021 from the Panel of Experts on Libya established pursuant to resolution 1973 (2011) addressed to the President of the Security Council," https://digitallibrary.un.org/record/3905159?ln=en accessed 05 July 2023.

[65] https://www.iai.co.il/p/harpy last accessed 5 July 2023.

[66] Rebecca Crootof argues that if we are to adapt existing regulatory regimes to cope with AWS successfully, we must recognize that we cannot persist in attempts to analogize them to either conventional weapons or any class of combatant currently recognized because none of these analogies captures all of the legally relevant characteristics of such systems, even if they may by turn help us answer one particular question or another (Crootof 2016, 56-58). She further argues that we limit our ability to consider all the forms automated weapons might take when we do so because focusing on AWS as either new types of weapons or replacements for human servicemembers ties us to the paradigm of weapons as physical artifacts (Crootof 2016). Autonomous weapons "need not be embodied at all" (Crootof 2016, 81).

outside of an operation, no operation takes place without a plan, no plan is written without a strategy, no strategy is selected without analysis, and none of these are chosen from among alternatives and implemented without properly authorized command.

Consider what is required when I say that, for the sake of discussion, we should assume that we will be able to create a sophisticated fully autonomous weapon system that has overcome any limitations on such a system's ability to function effectively as a technological matter or correctly as a legal and moral matter. Such a system might be one able to not only accurately identify a pre-selected target but also to select new targets independently. It would also be capable, in either case, of analyzing all the surrounding conditions to determine whether that target can be engaged at that time. That's a complex, multi-factored evaluation of the risk and acceptability of collateral damage—including civilian casualties—and the potential military benefit to be gained that calls for nuanced value judgments. The key standards delineating lawful employment of any weapons system, traditional or autonomous, are distinction (or discrimination) between military and civilian objectives, and proportionality (between military benefit gained from an attack, versus the negative impact on civilians and civilian objects). Both principles can be formulated as clear, simple rules of customary international law with little effort: 1) only military persons, places, and objects should be intentionally attacked, while civilian persons, places, and objects should be protected, and 2) civilian persons, places, and objects may only be endangered when the military advantage to be gained outweighs the risk. These assessments are in turn made according to standards like "reasonableness" (in assessing military necessity) and "excessiveness" or "feasibility" (in avoiding collateral damage).[68] Unsurprisingly, there are no objective definitions of these standards because it would be impossible to anticipate all possible circumstances, and LOAC makes no attempt to do so, instead mainly setting out a limited number of prohibitions – such as attack on civilians—and permitting all else.

[68] *See* GC Art. 57 para. 2 sub-para. a AP I; CCW Protocol III art. 1(5).

We can and do assign numerical values to anticipated levels of collateral civilian damage for use with algorithmic tools that help human beings make proportionality, including sophisticated software specifically designed to assist in making these calculations.[69] However, the subjective element currently remains; a human commander must decide whether a proposed action is permissible according to the computer-assisted analysis of the potential military advantage to be gained.[70]

Our hypothetical weapon system that is capable of legally and morally acceptable fully autonomous offensive function would have to be able to complete that same analysis—including some computerized substitute for the judgment calls about collateral damage and military necessity that are involved. That same analysis applies whether we're talking about the lowest-level ground troop trigger puller or the commander authorizing a remote (non-autonomous) drone strike. It also applies to commanders approving mission or operation plans at various levels of authority. At the highest levels it applies to broad strategic choices that can influence everything from how standing rules of engagement will be modified for a particular operation to which sorts of units will be deployed, what targets they will be authorized or forbidden to engage, and what weapons and tactics they will be authorized to employ. Those high-level officials also make choices about how changing domestic and international conditions should influence an overall national security strategy aimed at protecting important national interests. Someone must decide what those national interests are, what resources are available to protect and promote them, and how they should be applied to that end.[71]

Recall that we've accepted for the sake of argument that positions like Arkin's which claim that robot trigger pullers will perform better than human ones do are correct and that we can

---

[69] *See, e.g.*, Schmitt and Thurnher, 2013, at 254-5, discussing the U.S. CDEM; "Methodology for Combat Assessment" JCS Instruction CJCSI 3162.02.

[70] For example, the US DoD LoWM § 5.12 states that "Combatants must refrain from attacks in which the expected loss of civilian life, injury to civilians, and damage to civilian objects incidental to the attack would be excessive in relation to the concrete and direct military advantage expected to be gained," and that this evaluation "involves both professional military judgments as well as moral and ethical judgments evaluating the risks to human life" (§ 5.12.3).

[71] Some of those "resources" are human.

actually manufacture them. That would imply that they do an objectively better job at performing the seemingly-subjective analysis human agents are currently responsible for, just like my computer can perform complex mathematics better—meaning more quickly and more accurately—than I can, or even employ an algorithm to do an ostensibly better job deciding which of these sentences require commas and where.[72] My computer is clearly a better mathematician than I am, and perhaps it is, or could one day be, a better copy editor. Should we therefore conclude, though, that it can be a better author in every way that counts? Could it write "better" poetry, for example? It is a better mathematician already. There are arguably mathematical elements involved in certain poetic choices, and there are likely patterns in what we consider to be good poetry it could analyze and imitate. There are quantifiable elements and patterns involved in the military decision process as well. Given a sophisticated enough algorithm, a computer system should be able to do a better job at it just like it could theoretically do a better job deciding whether it had identified a correct target, and whether it ought to engage that target. Human agents can only collect, absorb, and analyze so much information at one time, and only so quickly. Computers can handle much more information, they can do it faster, and they can do it more accurately. If we had better ones than we do now, instead of using an algorithm only to assist in a collateral damage estimate, an evaluation of which proposed courses of action are most likely to succeed in a months-or-years-long operation plan, or track resource availability and allocation we could use them to make those decisions outright the same way we could use them to select and attack discrete targets.

Target selection at any level is, arguably, always a strategic decision. (Roff 2014). If we can trust robots to do it at the lowest levels, why not the highest? Conversely, if we could not justify automation at the highest levels, why would we accept it at the lowest? If a computer can better

[72] Anyone who has played with ChatGPT, Google Bard, or any other generative grammar system—or carefully audited the suggestions of Microsoft Editor should understand why this discussion requires that we assume *all* the current performance limitations so-called AI currently faces can be overcome.

perform the seemingly-subjective legal and moral analysis to decide whether a single physical target should be attacked with a single physical weapon at a single moment in time than a human agent can, it should be able to do a better job at all strategic decision making, from analyzing the entire network of national interests, national resources, and international relations that influence national strategy on down.[73] Commanders and their staff, especially planners and intelligence personnel, are responsible for gathering, understanding, and acting on vast quantities of information. They must often do so under great pressure and with limited time and resources. If we have reasons to think that autonomous weapons could better, or at least more accurately, perform the functions of conventional weapons and human combatants because they can gather, analyze and apply greater quantities of data more accurately, more quickly, and without emotional bias then we certainly have reason to believe that computerized systems could better assimilate the information available, identify and gather what is missing, and evaluate possible courses of action to identify which is "best" than can human commanders and their staff. Likewise, if we have reasons to think it is better to employ autonomous weapons than human combatants to relieve the burden of PTSD and other psychological effects that those combatants often face in the wake of decisions made on the battlefield—even when those decisions were correct by applicable legal and ethical standards—we should think we have just as much reason to relieve the psychological burdens their commanders bear as well. If we can make warfare more compliant with legal and moral principles, reduce unnecessary suffering, and make ourselves more likely to accomplish missions at the lowest level accurately and efficiently with replacing human agents with robots, imagine the advantages we could

[73] Heather Roff calls this the "Strategic Robot Problem." Roff, invoking the distinction between those weapons that are fully autonomous (able to employ learning and sensing technology to select targets independently), attributes strategic significance to the substitution of machine learning capability for human target selection decisions (213). In this discussion, I have argued that we ought to consider a larger class of decision making than Roff addresses, but her description of the problem shares the recognition that the relevant point of concern is the automation of what otherwise would be an important exercise of human military decision making.

see if we replaced their commanders? It is for this reason that I have argued we should look at the question of automation in the military not as a question about robot weapons, or even a question about robots or weapons at all. It's a question about replacing delegations of government authority to human agents—servicemembers, commanders, and even top-level officials—with computer systems that can process more information, more quickly, and more accurately. That should mean they are able to make objectively better decisions. If there are any reasons that we should not make such a replacement, that would indicate that we need something more from them than that they be objectively correct. The next section will argue that much like an important human element of expression would be missing from computer-written poetry, there is a necessary human element that would be missing in our hypothetical AI command structure, and which is necessary to justify our delegation of at least some classes of government decisions. The decision to take a single life in a single moment of combat is one example of such a decision. The decision to authorize that life and many others to be taken in a particular mission, operation or campaign is another, as is the decision that this national interest is important enough to justify the use of violent force but not that one or that military personnel ought to be deployed, that drones should be used instead, that we should send aid or equipment instead of personnel or that another state with which we have cooperation agreements should be pressured to take on the burden of an operation in our place.

Section IV. Automation, Delegation, and Trust in Military Decision Making

Alarm about AWS has more to do with the purpose or manner of its use (offense as opposed to defense, the application to which both its destructive capabilities and its autonomy are put) than any distinction in the lethality or non-lethality of the system itself or the degree of its autonomy. The substance of that distinction is one about what types of decisions are delegated and how much discretionary authority is abdicated to machines or retained by human beings. In the case of autonomous weapons, specifically, the focus is on the lower engagement-to-engagement level. In

the case of automation generally, embodied or not, the decisions that theoretically *could* be automated (assuming practical challenges are overcome) reach the highest levels of strategy and, therefore, of political and moral significance. This section will argue that the automation of military decision making should be evaluated as a delegation of authority.[74] To be justified in condoning the use of deadly force by a single human agent against a single human target, we must be able to trust that the authority will be used correctly. That in turn requires that the agent we have delegated such authority to must be trustworthy, a distinction I will discuss below. If these elements are necessary in decisions with such serious but limited scope, they are at least equally necessary in higher-level decisions to authorize any such uses of force—let alone many—in the first place. The first part of this section will address what type of trust is required. The second will explain why automated systems are not capable of satisfying that requirement.

A. Automation as Delegation

The distinction between weapons for defensive and offensive purposes highlights that it is not the question of lethal *capability* that raises objections. The distinction between "lethal" autonomous weapons ("killer robots" or even "slaughterbots")[75] has little to do with the destructive capacities of such weapons. Likewise, the issue does not materially turn on any *technical* distinction as to whether the system in question is "semi" or "fully" autonomous since systems like the C-RAM meet many of the proposed definitions of a "lethal autonomous weapon" in that they are equipped with rounds more than capable of ending a human life, and they are capable of detecting and engaging targets without human intervention in that process. The fact that a human operator defines

[74] It may be more accurate to describe it as a derogation or abdication in place of delegation, but for simplicity's sake I have continued to use the term delegation because the important point is that our analysis of whether a given human agent or some automated system can be permitted to make a given decision should ask whether the decision in question is one requiring trust and trustworthiness in a way machines cannot replicate, or not.

[75] *See, e.g.*, https://www.cnbc.com/2021/12/22/un-talks-to-ban-slaughterbots-collapsed-heres-why-that-matters.html accessed 3 July 2023.

targeting criteria before activating such a system does not change that the "decision" to launch a potentially deadly weapon at a given target has been delegated to an array of sensors and algorithms. The relevant difference as a question of ethical and political responsibility for the decision to employ force has to do with using such capability for defense instead of offense. The question is entirely one of employment—of just what targeting criteria that human operator programs into the system. That question is, in turn, a question of what orders have been issued (and what discretion granted) to that operator. Those orders, in their turn, are constrained by a political grant of authority to the military to engage in the use of force. The flow of discretion in interpreting how to use that political authority slows from a river to the tiniest of trickles by the time it reaches its endpoint, but that changes its volume, not its nature.

A meaningful assessment of whether there are limits on the permissibility of automated military decision making[76] requires an assessment of what matters about those decisions. We must consider whether the correctness of their substance is most important, or the merits of their procedure, and what makes a good military decision, or a good decision about the employment of state-sanctioned destructive force, a good one. In doing so, we should also wonder whether to prioritize accuracy above all else, as many arguments in favor of AWS do, or the reduction of presumably unnecessary suffering—physical or psychological—by human combatants. All these things matter, but none matter to the exclusion of all else. At least one of them, the possibility that we can eliminate the psychological burden of warfare on military members (as well-intentioned as it may be), should be rejected because it is at odds with an important requirement for the delegation of military decision-making authority. The sorts of military decisions at issue here are decisions about when, where, and how to employ force. The authority to engage in sanctioned violence is the

[76] Understood as the delegation of authority to computerized processes instead of human beings in any measure and at any level.

flagship power of the state, whether applied domestically in the form of criminal enforcement and punishment or internationally in the form of violent conflict. It is that flagship power of government that flows through the layers of military command to end with the sanctioning of violent actions taken by individual servicemembers so long as they are properly authorized and in compliance with the law. Automation of any part of that process transfers some of this power from human agents to computerized processes. It is, therefore, a significant decision about what constitutes the proper flow of the most important of all state authorities.

Among considerations for whether or how this adjustment to the flow of authority can be implemented is what can be automated without undermining the internal trust military leaders require to do their jobs effectively and the external trust other branches of government and the public at large must have in the military to maintain the legitimacy of its actions. Public trust in government decisions, including those made within the military, is crucial. The trust required here is different from simple reliance (expecting that the decisions made will be correct or accurate without that expectation being unduly disappointed) (Ryan 2020). It requires that we both be able and willing to place our trust in the decisions that will be made and the way they will be made—and that we are justified in doing so, despite the chaos and unpredictability of conflict. It requires that the decision-makers be trustworthy, meaning we are not mistaken in placing our trust in them (Hardin 2002, 30). We don't want them just to do the right thing. We want them to do things for the right reasons.

Those reasons, and the intention and motivation that they express, are one of the essential things that we lose when we automate. Intention and motivation matter because it may be difficult or impossible in some situations of conflict to endorse any course of action or outcome as unequivocally right. For another, we often accept that the intent behind a given use of force does alter the legal and moral character of its impact. In these cases, we resort to evaluating the legal and moral correctness of a decision by considering the responsible agent's motivation and intention. The

doctrine of double effect is directly incorporated into the often vague, subjective legal evaluations of proportionality, distinction, and military necessity. Our being able to rely on those intentions and motivations in a decision-maker is part of what makes it appropriate to delegate the life-altering and life-taking authority for state-sanctioned violence to that decision-maker. If they are absent or improper, we should not make that delegation whether we are talking about a human agent or an autonomous system because that agent or system would not be appropriately trustworthy. Military decisions must be trustworthy, such that public trust in them is warranted. That requires that they be made by agents who have the proper motivations and intentions. Part of having the proper motivations and intentions, in these cases, in turn requires agents who are both able and willing to assume responsibility, and the associated psychological burdens, for these kinds of high-stakes decisions. Their attitude towards that responsibility, and the reality of those burdens, shape their reasons-responsiveness in ways that even the most sophisticated autonomous systems cannot emulate. Those attitudes are pivotal for both the moral and legal evaluation of which uses of force are justified, and which are not.

B. Automation and Trust

The appeal to the Martens clause as a basis for banning AWS/LAWS may lack the specificity and concreteness to carry compelling legal force, but it does highlight an important locus of popular concern and an accompanying threat to public trust in military decision making. That concern is a crucial hurdle for any effort to delegate what would otherwise be government decision authority to a computerized system while maintaining ethical and political legitimacy. Whether automation will cause us to lose our ability to demand transparency from decision-makers and hold them accountable for their transgressions and failures is one aspect of that threat that is captured in the literature on responsibility gaps: Most of the literature on addresses such gaps in our ability to *assign* responsibility for automated acts (Santoni de Sio & Mecacci 2021 at 1065-1066). However, there is

another side to the responsibility question at work when it comes to trust and trustworthiness requiring a decision-maker that is able to *assume* responsibility.

There is a difference between whether we can rely on a system to perform correctly, and whether we are justified in placing our trust in that system—whether the system is trustworthy. (Ryan 2020, 2752). When I suggested that we assume obstacles to creating effective, accurate, truly autonomous systems could be overcome, and that those systems would in fact be objectively better at their tasks than human agents are, I was asking that we assume they would be reliable in the sense that Ryan uses the term. We would be able to *rely* on them to act in ways that are effective and objectively correct. Relying on them, however, is not the same as trusting them and reliability is not the same as trustworthiness (Ryan 2020, 2757-2758). For example, we might be able to rely on the threat of punishment to cause someone who lacks the appropriate moral motivations to nonetheless refrain from causing intentional physical injury to other people, or willfully destroying their personal property…so long as they believe they will be caught.[77] That is not the same as our being able to trust them not do those things. In fact, our belief that external motivation for their actions is the thing that causes them to behave correctly specifically implies that we do not trust them and that they are not trustworthy, because we know they lack the proper internal motivations. Discussions about assigning responsibility address these sorts of external motivations. They are much more about reliability, whether we're discussing autonomous systems or human agents, than they are trust.

Justified trust, and therefore trustworthiness, depend on internal motivation. Theories of trust that truly deal with trust and not reliability, Ryan (2020) argues, are those that place importance on our being justified in expecting that the trustee is affected by our having trusted them, meaning they recognize the responsibility that accompanies our trust (affective accounts), and that they are

[77] Ryan cites an example of a supervisor who can be relied upon to refrain from sexually harassing employees because he will face consequences if he does not (2758).

internally motivated to fulfill the normative obligations that accompany our expectations and our trust (normative accounts).[78] The trustee is motivated by their having been trusted, and they are motivated to satisfy the demands of that trust. This response requires psychological capabilities that no AI, no matter how sophisticated, can possess (Ryan 2760-2761). One way to look at this aspect of trust is to say that, to be trustworthy, the trustee must assume responsibility for the trust that has been placed in them and they must be responsive to our expectations about how they will fulfill their corresponding obligations. When we trust someone in this way, we believe that they have the right motivations (Faulkner 2017, 109) or the right attitude towards the trust granted to them.[79] Automated systems are not capable of having any motivations or attitudes at all, let alone "the right" ones, but this is the trust component that is necessary for our delegating authority over the use of violent force to any decision-maker.

This is important for two reasons. First, as I've noted above, military decision making requires trustworthy agents because the applicable legal and moral standards often depend on the same internal components of trustworthiness that distinguish it from reliability. Second, we should not delegate where we cannot trust, especially in areas like military decision making where the stakes are high, the impact of many decisions is irreversible, and the transparency on which external assurances of reliability depend is lacking. No system of authority can be trustworthy if that authority is delegated to agents who are not themselves trustworthy. No government can maintain a

[78] Ryan considers each of three varieties of trust theories to reach this conclusion: rational, affective, and normative accounts. Ryan concludes that while AI may be able to fulfill rational accounts of trust, meaning ones that describe trust as our ability to predict whether the object of our trust will behave in the rational and expected way in a given situation regardless of motivation, these accounts do not properly describe trust at all because they ignore most of the interpersonal significance of the phenomena in favor of a focus on results. They are instead accounts of "reliability." (Ryan 2020, 2759; Faulkner 2017). The other two accounts include criteria AI is unable to meet, which means that although it may be reliable, AI cannot be trustworthy: if we place trust in it, as opposed to relying on it, that trust is misplaced. Affective accounts of trust, Ryan says, place importance on our being able to expect that the trustee is affected by our having trusted them in the appropriate way. They should recognize the responsibility our trust, and its associated expectations, creates and wish to live up to those expectations. (Ryan 2020, 2760).

[79] Faulkner describes trust, as distinct from reliance alone, as bestowing our confidence not only in how we expect another person to act but also in their character and motivations ("to trust someone is to take an optimistic view of that person and their motivations").

healthy relationship with its citizens or the international community if it, and its authorized decision-makers, are not worthy of trust.[80]

These requirements rule *out* automated systems as proper delegees for at least some types or levels of government authority because of the way they cannot respond to that delegation that human delegees can: they cannot assume responsibility for their actions. They cannot have the proper attitude towards those decisions because they cannot experience the burden of deciding to employ force that weights lives against interests, or one life over another life. They cannot appreciate the burden which remains—if an agent possesses the appropriate attitudes and motivations—regardless of whether that decision was legally and morally correct or not. The legal and moral restrictions we place on the use of violent force recognize that it is an undesirable last resort. We expect that nations will employ it only when and as much as necessity demands. Anything more violates those restrictions, and the fact that any violence at all was necessary ought to occasion feelings of regret. In the end, that regret indicates whether a delegee has in fact assumed responsibility. Its presence can tell us whether they have approached such authority with the proper attitude and the capacity to experience it motivates that delegee to use their authority correctly in a way that makes them trustworthy instead of only reliable. It is similar to what Bernard Williams refers to as agent-regret in the context of moral luck, as opposed to bystander regret: regret centered around the wish that we had acted differently not because we made a morally objectionable choice but because some otherwise un-objectionable choice that we made had regrettable results. (Williams

[80] Margaret Urban Walker describes trust—and its restoration after wrongdoing—as an essential component of moral relationships she identifies as necessary to our ability to maintain social relationships of any kind, whether strictly interpersonal or with a larger community or institutions (Walker 2006). Walker's account of moral relationships combines elements of what Ryan describes as normative and affective trust paradigms and helps illustrate why it is not enough that we delegate decision making to reliable systems. We have confidence, she says, that some authoritative, shared, and worthwhile set of standards for conduct exists. We trust that others will be responsive to these standards in the right way, and we hope that we have not misplaced that trust (that the individuals and institutions in which we place it are trustworthy) (24). There remains an element of reliance to this sort of interpersonal trust. What distinguishes from the accounts Ryan dismisses as reliance alone instead of trust (or trustworthiness) are the contents of the expectations we believe will be fulfilled.

1977, 27-28). Agent-regret is something more than the feelings an equally non-culpable bystander might have of being sorry that something tragic has happened. (Id.) It acknowledges that even if we could not have acted differently or did not act in a way that others would deem blameworthy, we have been the agents of a regrettable outcome. The decisions military decision-makers face are different than Williams' example of a "[truck] driver who, through no fault of his own, runs over a child" in that they lack the involuntary character of a tragic accident, but the dissonance between moral and legal culpability on the one hand, and regret born from internal feelings of responsibility, on the other, is similar. Agent-regret can be an ugly, painful feeling. It forces us to recognize the inescapable consequences of our actions, even those that are not blameworthy. It forces us to take our roles as agents seriously. No amount of bystander regret over automated violence can motivate trustworthy military decision making the way that agent-regret over agent-authorized violence can. This is important because these decisions are often ones where no available answer or outcome is a wholly good one. There will often be regrettable consequences no matter how well the decision that caused them complied with the legal and moral guidelines for warfare. We should not make them too freely or easily. The psychological burden of agent-regret in these cases is a normatively important one; the agents in question would not be trustworthy without it, no matter how reliable they might be.

III. The Lonely Moment: Why Fallible Human Leaders are Required

In the previous section, I argued that it is the psychological capacities of human agents, and the associated psychological burdens, that make them trustworthy military reasoners in a way that autonomous systems cannot be, even if we have assumed that those systems may be more reliable. I have also argued that it is trustworthiness that justifies delegation, not reliability. The implicit claim is that we should prefer to trust an important type of government authority to the set of potential decision-makers that is more, not less, fallible—a position that this section will defend.

The underlying reason that we should not be too quick to delegate decisions, whether high or low stakes, to autonomous systems, even if they were proven able to make them more accurately, quickly, and efficiently than human beings is the same reason we require individual, potentially fallible leaders within any system of authority, laws, and institutions. It is the reason that Plato, thousands of years before autonomous weapons or artificial intelligence were imagined, called written law a poor substitute for expert leadership, sure to still require a skilled and presumably trustworthy individual to fill the gaps even the best body of law will unavoidably leave. (Plato, trans. Rowe 1999). Any decision made on behalf of the state by virtue of some delegation of authority is one small part of the business of governance. That business is a human business. We talk of governing states, but we mean governing people. We talk of governments, but we mean a set of laws and institutions created by people and the decisions made within those institutions and in accordance with those laws by people. The decisions a government makes are important ones, and many of them—such as the decision to employ violent force in warfare—are irreversible. It would not be reasonable to grant authority to make such decisions where we cannot also bestow our trust. It would not be responsible to do so without assurance that that trust has not been misplaced. To deserve and *maintain* public trust, the government must be able to show that it is trustworthy. If true trust and trustworthiness require the affective and normative elements Ryan, Faulkner, and others argue they do, this means that while we may rely on laws and intuitions to channel and influence the use of government authority in acceptable ways, we must be able to place our *trust* in the people who make decisions under that authority.

In the case of the military, that means that decision-makers, especially leaders, should show themselves trustworthy. They might do this by making themselves publicly accountable through the sorts of transparency measures expected of most government decision making, but in many of the most important cases, true transparency is impossible because of the need for security—to provide

security—secrecy. All trustworthy leaders should be willing to *assume* the burden of responsibility for the decisions entrusted to them rather than merely making themselves vulnerable to its assignment in the event of bad outcomes, but this need is at its greatest in the context of low transparency and high stakes military leaders face. For any trustworthy leader,[81] there is a lonely moment of decision where, although she may trust in the orders and advice of others, she must personally accept the burden of responsibility for the act she carries out in the name of the state. There is often no fully right course of action and no fully desirable outcome available in conflict. We should not make the inevitable damage easier by allowing automation to erase or reduce the immediacy of the fact that innocent people will suffer, that even those who are not innocent per se are still human beings capable of suffering.

If we accept the position that trust and trustworthiness have to do with a trustee's being motivated to live up to a set of norms, we can view a military leader's assumption of risk (for bad decisions), reliance on other similarly-situated peer leaders or subordinate leaders, and efforts at transparency and active accountability as evidence of her both exhibiting trust and trustworthiness in this way both horizontally and vertically through the channels of dispersed state authority. She can trust the recommendations and decisions of her peers, subordinates, and superiors because she trusts them to be responsive to the high stakes and normative impacts of the activity they mutually engage in in the appropriate way. Other institutions can likewise trust that military decision-makers will wield their authority responsibly for these reasons. The members of the public can likewise trust (at least more than they otherwise might) that the people within government institutions who have been granted the authority to act without direct popular oversight will do so in ways consistent with broadly accepted norms and the rule of law. This network of trust relationships is possible because

[81] In this sense, any person entrusted with the final discretion over an instance of force should be considered a "leader," because in that moment she embodies the most fundamental exclusive authority of the state.

everyone involved in making these decisions shares an attitude of agency towards the process. They have the capacity to experience agent-regret for the consequences of their decisions and treat them with the seriousness that they deserve; they all have some moral skin in the game.

This personal moral investment makes military decision-makers responsive to the risks involved (the good or bad consequences of any particular action or inaction). It will likely result in morally better decisions even though everyone involved faces epistemic and reasoning-ability limitations that automation might overcome. Even the best machine reasoners cannot replicate this trustworthiness of decision-makers because a machine cannot be responsive to moral risk and psychological burden in the same way. No computer will have to "live with" the consequences of a careless targeting decision or excessive hesitation in authorizing an attack during the chaos of an evolving kinetic situation. Some scholars think that this could be an advantage of AWS (Arkin 2010, for example, highlights the occurrence of PTSD even among drone operators as a bad effect of human engagement in warfare that the use of AI could ameliorate, a perspective echoed in the broader context by Munch et al. 2023, arguing that in some cases responsibility gaps are desirable because they may alleviate the psychological burdens of decision making). The desire to spare human decision-makers from the burden of these morally risky decisions is well-intentioned but misguided.[82] The state monopoly on the use of violent force is a fundamental feature of government authority. The responsibilities accompanying its delegation to military leaders should not be taken lightly, even for the sake of accuracy in decision making. There is a difference in kind between the dispersal of that authority among even a vast number of appropriately norm-responsive decision-makers who must accept the burdens of their decisions along with the power they have been granted and its dispersal to even the most sophisticated of autonomous systems that are

[82] My objection to the claim that automation is desirable because it spares human beings from the heavy psychological burdens of warfare is in no way intended to imply that the suffering they cause is deserved, desirable, or not worthy of our attention and concern—only that automating conflict is not a proper solution.

incapable of being responsive in the relevant way because they lack the capacity to assume these burdens or feel their weight.

I previously mentioned the problem of responsibility gaps sometimes raised as an objection to the use of AWS as well as to other uses of AI. Responsibility gaps exist when it is unclear to whom we ought to assign responsibility for the actions of AI or where the control of any human being over the action in question is so attenuated that struggle to assign any responsibility at all. Objections that AWS should not be used because they cause responsibility gaps were among the objections suspended during this discussion. That is, I assumed there could be some way of closing all these gaps: that we could create and use AWS without ever finding ourselves in a situation where no one could be blamed for their mistakes. To complete the argument as to why interpersonal trust among military decision-makers is 1) something that is necessary and that cannot be replaced by automation, no matter how objectively superior that automation may be, and 2) necessary in turn for maintaining public trust in military decision making I will now revisit the notion. There is a relevant responsibility gap here, but it does not have to do with our ability to assign responsibility. Instead, it has to do with the ability and *willingness* to assume it, along with the associated psychological burdens.

Accounts of trust and the moral relationships that require it (Walker 2006, Jones 2017) place great weight on our entitlement to call to account those who fail to live up to our normative expectations of them. However, such recriminations are of little value if they can receive no meaningful response. This is why I have said that if we automate certain high-stakes decisions about which we would wish to make strong recriminations should they betray our trust, there is a responsibility gap at work that is different from those having to do with *assigning* responsibility alone. The existence of this sort of gap at work when the reduction or avoidance of the psychological

burdens of warfare are cited as an advantage of AWS.[83] It is no such thing. It is exactly their awareness of the risks they take—and the bad consequences that may result—that ensures trustworthy military decision-makers will use their authority judiciously. Suppose we were to replace a brigade commander's staff with sophisticated computer systems that use AI to plan operations for its component units.[84] The responsibility that the commander feels for ordering her troops to follow the AI-generated plans should motivate her to carefully review its proposals, and is categorically different than an investigating officer's ability to later assign responsibility for a tragic outcome to that system, the team of programmers who created it, or the government contracting authority who procured it. Suppose another sophisticated computer system exists that could perform the commander's review of the proposed plans more quickly and accurately. If we were to use it, we could retain the second sort of responsibility but at each subsequent level where we prioritize reliability over trustworthiness and desire to automate, we lose more of the former.

Placing priority on that former element of assumed, agent-responsibility instead of assigned responsibility alone would change our analysis of whether these sort of automation are acceptable, and if so how pervasively we can use them before that sense of assumed- or agent-responsibility has stretched too thin, and along with it our ability for agent-regret to act as a moderating influence on the desire for accuracy, speed, and efficacy at all costs that often lead us to automate in the first place. There are some government authorities that should be delegated only to decision-makers who are trustworthy in a way that requires assumed responsibility and the capacity for agent-regret, which we might call "inherently human functions." As we introduce increasingly sophisticated automation into the processes of government in any area, we should be careful to identify and then reserve inherently human functions for human agents in the same way we reserve inherently governmental

[83] Munch et al. (2023) describe such gaps as a possible societal benefit of automation through AI generally.
[84] In the United States Army, a brigade is a unit of about 3000-5000 soldiers, organized into battalions of about 1000, which are in turn organized into companies of usually 200 or fewer soldiers.

functions to government agents. The argument over whether we can ever properly allow an autonomous system (one with no human-in-the-loop) to independently apply lethal force, for example, asks whether the authority to apply such force under the banner of state authority is or is not such an inherently human function. It asks whether there are any important features of that decision-making process that cannot be retained by an autonomous system that performs with at minimum the same level of correctness in its compliance with legal and moral restrictions on the use of force as a human being does. In some instances, it further asks whether we are obligated to rely on machines instead of human agents if it is the case that their decision making would be *more* correct. It asks, that is, whether correctness—which could be satisfied by reliability alone—is all that matters.

Because the problem of AWS and the delegation of lethal authority to machines instead of human agents asks that question about only one of many delegations requiring us to know when trustworthiness is required, and when reliability will instead suffice, we need to answer it in a way that will allow consistency across all such questions. That calls for criteria to tell us which functions are inherently human, and which are not. The effort to define functions that are similarly inherently governmental, and which therefore must be delegated only to federal employees who are subject to federal oversight and presumably worthy of public trust, is not a bad place to start. The term is defined in section 5 of the Federal Activities Inventory Reform Act, Public Law 105-270 as "a function that is so intimately related to the public interest as to require performance by Federal Government employees," which includes "functions that require either the exercise of discretion in applying Federal Government authority or the making of value judgments in making decisions for the Federal Government," as well as the authority to "bind the United States to take or not to take some action by … authorization, order, or otherwise," to "determine, protect, and advance United States economic, political, territorial, property, or other interests by military or diplomatic action,"

and to "significantly affect the life, liberty, or property of private persons." Those decisions considered too "intimately related to the public interest" to allow their performance by contractors are likely to also be those we cannot automate away any more than we can contract them away, at least not fully, without undermining public trust. This does not mean we cannot use the powers of automation to *enhance* our performance of such functions, however. For example, I stated above that a view of military automation limited only to robot weapons misses the greater potential it could have for information gathering and analysis or for planning and proposed we consider the differences in one example of a commander who reviews and approves of plans generated by such an automated system as a contrast to our having replaced that commander with yet another automated computer system. It may be that the first option is acceptable if we consider the review and approval of such a plan to be the inherently human function at work. The provisions delineating some functions inherently governmental provide a parallel exception for "gathering information for or providing advice, opinions, recommendations, or ideas to Federal Government officials." There will be come functions so inherently human that they may never be automated, just like some functions are so inherently governmental they can never be contracted. Others will be appropriate for exceptions like the one noted above, and proper for automation. Many activities, such as the process through which plans are created, reviewed, and approved, will exhibit a mixture of both. In those cases, a semi-autonomous or human-in-the-loop approach may still be permissible. Reframing the problem in this way provides a starting point for making the conversation about our very real legal and moral worries about automation into a conversation about consistent, workable policy founded in legal and moral principles instead.

I have argued in this paper that we should view the problems presented by AWS not as a unique problem about either robots or weapons, but as one aspect of the problem of automated military decision making. The problem of automation in military decision making, in turn, is one

manifestation of a yet larger problem regarding government automation, and that problem is one that has to do with the flow of government authority in general: it is a problem of delegation. Questions about AWS/LAWS, then, are all under the surface questions about whether a certain type of delegation of authority may be properly delegated to non-human "agents," and that question is an important one for us to answer about any use of automation in government—especially since the arguments in favor of its use for AWS/LAWS apply with equal force to nearly all military decision making. To ask only about whether or when we may permit "killer robots" is to ask about only one small tree within a vast and thorny forest. Although Asaro's appeal to "human dignity" as an argument against the moral permissibility of LAWS is vague, its insistence that something necessary that is lost when we automate the taking of a human life has to do with the agent-side recognition for the seriousness of the act. These decisions matter to human beings in a way that goes beyond legal or moral principles designating them lawful or unlawful, justified or unjustified. Even when a destructive act is justified, we ought not to feel entirely good about it. We ought not to be unburned by feelings of responsibility for the bad character of even the least bad outcomes. I do not hereby argue that no degree of automation is appropriate, only that we should implement such technologies in ways that respect their status as repositories of important state authority, and which do not undermine trust and trustworthiness within that necessarily human process.[85]

[85] Cavalcante Siebert, 2023 approaches a similar analysis from the perspective of when and how "meaningful human control" ought to be part of AU design to address the problem of assignment-side responsibility gaps.

Works Cited

Anderson, M and Anderson, L.A. (2015). Toward Ensuring Ethical Behavior from Autonomous Systems: A Case-Supported Principle-Based Paradigm. *Artificial Intelligence and Ethics: Papers from the 2015 AAAI Workshop*

Arkin, R. (2008). Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture (Technical Report GIT-GVU-07-11) https://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf

--(2010). The Case for Ethical Autonomy in Unmanned Systems. *Journal of Military Ethics, 9*(4), 332-341

Arkin, R., Ulam, P., and Wagner, A. (2012). Moral Decision Making in Autonomous Systems: Enforcement, Moral Emotions, Dignity, Trust, and Deception. *Proceedings of the IEEE, 100*(3), 571-589

Asaro, P. (2013). On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making. *International Review of the Red Cross, 94*(886), 687-709

--(2020). Autonomous Weapons and the Ethics of AI. In Liao, S. M. (Ed.). *Ethics of Artificial Intelligence*. New York, NY: Oxford University Press

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. B*ig Data & Society January–June 2016*: 1–12

Cherry and Rizzotti https://lieber.westpoint.edu/understanding-self-defense-law-armed-conflict/

Coeckelbergh, M. (2012). Can We Trust Robots?. *Ethics and Information Technology*, *14*(1), 53–60

Crootof, R. (2018). Autonomous Weapon Systems and the Limits of Analogy. *Harvard National Security Journal, 51*, 51-83.

Cummings, M. L. (2019). Lethal Autonomous Weapons: Meaningful Human Control or Meaningful Human Certification? *IEEE Technology and Society Magazine, December 2019*

Faulkner, P. (2017). The Problem of Trust, in Faulkner and Simpson (Eds.). *The Philosophy of Trust*. Oxford: Oxford University Press

Hardin, R. (2002). *Trust and Trustworthiness*. New York, NY: Russell Sage Foundation

Horowitz, M. (2016a). The Ethics & Morality of Robotic Warfare: Assessing the Debate over Autonomous Weapons. *Dædalus, the Journal of the American Academy of Arts & Sciences*, *145*(4), 25-36

Horowitz, M. (2016b). Public Opinion and the Politics of the Killer Robots Debate. *Research and*

Politics, January 2018, 1-8

Jones, K. (2017). But I Was Counting On You!, in Faulkner and Simpson (Eds.). *The Philosophy of Trust*. Oxford: Oxford University Press

--(2019). Trust, Distrust, and Affective Looping. *Philosophical Studies, 176*(4), 955-968

Munch, L., Mainz, J. and Bjerring, J.C. (2023). The value of responsibility gaps in algorithmic decision-making. *Ethics and Information Technology* 25:21

Plato (1999). *Statesman*, translated by Christopher J. Rowe, Hackett Publishing Company, Indianapolis

Prakken, H. (2017). On the Problem of Making Autonomous Vehicles Conform to Traffic Law. *Artificial Intelligence and the Law, 25*, 341-363

Railton, P. (2020). Ethical Learning, Natural and Artificial. In Liao, S. M. (Ed.). *Ethics of Artificial Intelligence*. New York, NY: Oxford University Press

Roff, H. and Danks, D. (2018). "Trust but Verify": The Difficulty of Trusting Autonomous Weapons Systems. *Journal of Military Ethics, 17*(1), 2-20

Ryan, M. (2020). In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Science and Engineering Ethics, 26,* 2749-2767

Santoni de Sio, F. and Mecacci, G. (2021). Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them. *Philosophy & Technology, 34,* 1057-1084.

Cavalcante Siebert, L., Lupetti, M.L., Aizenberg, E. *et al.* (2023). Meaningful human control: actionable properties for AI system development. *AI Ethics 3*, 241–255

Scharre, P. And Horowitz, M. (2015). An Introduction to Autonomy in Weapon Systems, working paper. *Washington, D.C.: Center for a New American Security*, 7

Schmitt, M.N. and Thurnher, J.S. (2013). "Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict. *Harvard National Security Journal, 4*, 231-281.

United Nations Security Council. (2021). Letter dated 8 March 2021 from the Panel of Experts on Libya established pursuant to resolution 1973 (2011) addressed to the President of the Security Council. https://digitallibrary.un.org/record/3905159?ln=en accessed 05 July 2023.

Walker, M.U. (2006). *Moral Repair: Reconstructing Moral Relations After Wrongdoing*, Cambridge: Cambridge University Press.

Williams, Bernard. (1982). Moral Luck. In *Moral Luck, Philosophical Papers 1973-1982*. New York, NY: Cambridge University Press