THE LIMITS OF CONSEQUENTIALISM IN POLICY EVALUATION

Walter D. Valdivia

Abstract

Policy evaluation has been dominated by the logic of instrumental reason and the moral theory of consequentialism. When evaluation privileges instrumental reason, it excludes justifications and motivations that fall outside of that logic, particularly reasons of stakeholders traditionally excluded from policy debate. I argue that this weakness in policy evaluation can be redressed, and contrary to contributing to the decline of trust in government, evaluation can be a steward of democratic values in policy making.

1. Introduction.

The following discussion is motivated by a practical problem: the decline of public trust in government. Political scientists have identified this problem since the early 1970s using half a century of electoral survey data that included the question: “How much of the time do you think you can trust the government in Washington to do what is right?” (Alford, 2001; Citrin, 1974; Lipset and Schneier 1983, 1987; Miller, 1974; Nye, Zelikow, and King, 1997). The many putative causes for the public’s decline of confidence include heightened social tensions (e.g. around civil rights), domestic economic problems (e.g. unemployment, unmanageable fiscal deficits, and competitiveness crisis), widening of economic disadvantage (Bartels, 1

---

1 This question is from the American National Election Studies (ANES) produced by the Center for Political Studies and distributed by the Inter-University Consortium for Political and Social Research at the University of Michigan. Other questions in ANES are also used to compute the American Political Trust Index which does not differ from the specific question on trust mentioned above.
2008), or eroding national prestige (e.g. the apparent loss of U.S. clout in international affairs) (Nye, 2004a; 2004b, 2010), as well as incidents of political scandal (Watergate marked the beginning of the decline) and publicized failures of leadership and responsiveness. These disappointing realities are not by themselves the source of distrust because disillusion results from our perception of reality falling short of our normative expectations. The lack of trust in government is thus born from contrasting those perceived realities against notions of what government ought to be doing; which is why people in both ideological positions of U.S. politics are losing their confidence, the government does either too much or not enough (see Figure 2.8 in Alford, 2001). It is then important to recognize that the problem of public trust in government depends on the ways in which people form normative views of government, and the extent to which they causally attribute specific aspects of their perceived reality to government. Therefore, not only contemporary events of political and social life have driven the citizens’ confidence in government down, but also the ways in which normative expectations about government are formed, ways that in many respects are governed by a “consequentialist morality” and guided by “instrumental reason.”

This is the first link between the motivation and the problem of this inquiry. Instrumental reason is implicated as much in the decline of the public trust in government as it is in the foibles of conventional policy evaluation.² My goal here is to argue the latter point, that is, to examine deductively the limitations imposed by instrumental reason on the discipline of policy evaluation in light of the democratic

² The critique of instrumental reason has been argued famously by Max Horkheimer (2004|1974) and elaborated further by the Frankfurt School to what amounts to a critique of “consequentialism” (Horkheimer and Adorno, 1972; Marcuse, 1968; and second generation Habermas 1975, 1984, 1987).
commitments of this discipline. My formulation takes issue with the moral theory of consequentialism.

A second manner in which the parallel problems of trust in government and incomplete policy evaluation are connected is the citizenry’s exclusion from the policy making process. Over the last four decades we have seen a public confidence in government drop, but we have also seen increasing civic mobilization and activisms to include the lay public input in policy decisions (Inglehart, 2001), including areas traditionally insular from lay opinion such as those involving highly technical expertise. Notable examples are nuclear power (Kasperson et al. 2008), ecological degradation (Hajer, 1995), and biotechnology in the food supply (Winickoff et al., 2005) where groups of organized citizens have increasingly demanded greater inclusion in bureaucratic regulation and legislative policy-making.

In response to this situation, government has sought to increase spaces for wider participation in the policy making process (Kettl, 2000; Pollit and Bouckaert, 2004) and scholars have turned more eyes to public participation and engagement (Skocpol and Fiorina, 2000; Verba, Schlozman, and Brady, 1995; see reviews in Carpini, Cook, and Jacobs, 2004 and Lowndes, Pratchett, and Stoker 2001a; 2001b). Policy evaluation has not been impervious to this evolution in political relations. In fact, evaluators have begun to see their role in organizational learning as derivative from their principal concern with organizational outcomes (Fleischer and Christie, 2009) which may in part be a response to previous calls for involving stakeholders early and often in the evaluation process (e.g. Reineke, 1991). Nevertheless, I argue that these efforts will not be sufficient to correct the neglect of key public values in policy evaluation if the discipline does not revisit its own commitments to democracy. The
problem resides in conceiving of policy evaluation as an extra-political enterprise. When politics is brought back into our conception of policy evaluation, the questions of political representation and legitimacy re-appear. To address these questions I draw from two theorists (Brown, 2009; Catlaw, 2007) who reformulate representation and legitimacy in relation to democracy, expertise, and the bureaucracy, and deploy their proposals to the case of policy evaluation.

The argument proceeds in the following manner. In a deductive fashion, I first elucidate the problem of policy evaluation in relation to consequentialism. Then I revisit the concepts of “representation” and “legitimacy” to re-conceive the methodology of evaluation.

2. Consequentialism and policy evaluation.

Policy evaluation, for all its sophistication, remains underdeveloped in important areas. Conventional evaluation provides tools to measure the impacts of public programs and policies but it does not give any criteria to evaluate the value of those impacts. Short of the latter, policy evaluation is only a dull scale that tells us if we have more or less, but not if collectively we are better or worse. My aim with the observations that follow is to make a case for expanding the scope of policy evaluation and pathway for developing the necessary methodology.

The so-called “evaluation problem” is the problem of measuring policy impacts. The policy impact is the effect a policy treatment has on a social unit or entity such as a nation-state, an organization, an individual, etc. However, no entity can experience two states, treatment and no treatment, at the same time; hence, the
problem is to infer its “alternative state,” ex-treatment. Policy evaluation methods have focused therefore on the production of that alternative state and the attending problems, including the unit-homogeneity assumption necessary for inference, the ceteris paribus assumption needed to preclude misattribution of causality, and the effect-stability assumption that informs future public action (for reviews see Heckman, 2007; 2001; Weiss, 1998).

It is tautological to say that consequences matter; hence, the methodological focus of policy evaluation on consequences is self-evident and justified. However, it is also tautological that evaluation encompasses the assessment of value of the consequences of policies. In general, evaluations have taken two strategies to deal with the latter normative question: either they leave the reader to judge whether the direction and size of the impacts are desirable or they incorporate, in the evaluation methodology, an implicit normative theory of value. “Let the facts speak for themselves” say evaluators who think hardly anyone could argue with the value they attribute to the impacts they just inferred, or because they are confident that the moral system underlying their technical approach is fair. This is of course not altogether objectionable. For instance, if the impact of policy is fewer dead people, hardly anyone will deny that such an outcome is desirable. In contrast, when the impact is less well determined, for instance the value of a program to create jobs may depend on many factors including whether the new source of employment is sustainable and if it is not displacing other more sustainable sources of employment.

---

3 When the alternative state is inferred from historical time series of the variable in question, it is called counterfactual. We could use this term more expansively, for instance to include the control group in an experimental/quasi-experimental evaluation design, since this group is used to see what happens if the “fact” of treatment did not occurred on an identical individual, that is to say, the “counter the fact” state.
This is one area where evaluation methodology needs further development, so that we can move away from an intuitive valuation towards a more systematic appraisal of the value of outcomes.

The attention to consequences is not disingenuous. To start, there is a tradition in policy evaluation going back to Dahl (1963)\(^4\) and Gamson (1968) that prescribes considering those outcomes that can be tied to the intentions that motivated the policy, generally the program’s stated goals. That is a convenient strategy from a methodological and a political point of view. Confining attention to a well-defined set of effects determines the space of outcomes to be examined, which in turn determines, per force of convention, the space of causal relations.\(^5\) Otherwise, unabridged evaluation leaves the problem open to indiscriminate inclusion of effects and ad hoc causal relations. Exclusive attention to stated goals could also give a political advantage; it justifies a narrow focus of evaluation on the presumption that those goals emerged from a democratically legitimate policy process. Likewise, a public official may claim the same to fend off criticism on other less salutary consequences of the same policy. But criticisms cannot be dismissed too easily when ingenious evaluators uncover unintended consequences of clear and urgent concern.

Whether intended or unintended consequences, the measurement of outcomes, and the sizing of impacts have dominated the methodological efforts of policy evaluation. Evidently there is an underlying commitment to consequentialism.

\(^5\) It is of course possible to conceive infinite putative causal relations for a single effect. Nonetheless, disciplinary tradition and convention often allows but a few causal relations to be deemed plausible.
Elizabeth Anderson defines it as a theory of moral action satisfying the following conditions:

“First, it gives people the sole ultimate aim of maximizing intrinsic value. Second, it holds the fundamental object of intrinsic value to be the state of affairs. It assesses the value of a state of affairs independent of the values of persons, actions, motives, norms, practices, states of character, or anything else. Third, it assesses the values of these other sorts of things, or at least actions, rules, or practices, solely in terms of their consequences, broadly construed—that is, in terms of how effectively they bring about or embody the best states of affair... [F]ourth: that all values are ‘agent–neutral’. A value is agent-neutral if it gives everyone a reason to value it. A value is agent–relative if it gives only some people a reason to value it” (Anderson, 1993, pp. 30-31).  

Consequentialist normative theories such as utilitarianism and welfarism (Sen, 1979) have in fact lent policy evaluation a normative core. A case in point is cost-benefit analysis that combines methodological individualism with a normative theory of value such that individual valuation is mapped onto a single scalar measure of value (willingness to pay) rendering, in this way, individual values additive and social value measurable. Cost-benefit evaluation provides a straightforward representation of consequences—that is, of the policy impact—as a welfare gain or loss (formally consumer and producer surplus). At the same time, the change in welfare represents the social valuation and thus social value of the policy. The

---

6 In her own analysis, Anderson’s (1993) does not presume “agent-neutral evaluative consequentialism” and therefore, does not use the fourth condition.
linchpin of this methodology is evidently the premise that individual valuation can be mapped onto a single scale of measure.

It should be clarified why we link cost-benefit analysis and “willingness to pay” because much of standard practice in cost-benefit does not seem to derive common costs or benefits from willingness to pay. The unifying method of cost-benefit analysis is to transpose every consideration of analysis onto a single scale, which is often expressed in a monetary unit (at a given point in time), so that a net calculation of pros and cons can be performed (and time-preference accounted for).

Take for example, the regular cost-benefit analysis practice of monetizing the cost of a social program, not measuring willingness to pay but taking the realistic approach of adding all public expenditures associated to such a program. This is surely an acceptable approach. Still, the cost of different levels of service may not be appropriately estimated simply at the unit cost of the level actually observed. It is not unusual that the marginal cost or cost of one additional unit of service may come at a cost higher than the last unit serviced, as is the case of a traditionally upward sloped supply curve. How are those other levels of service—in the vicinity of the observed level—estimated? The answer is given by approximating the public expenditure of the marginal unit via willingness to pay. The same may happen for the demand side, as the last consumer may be the last one willing to pay that high a price for the service. This is hard to imagine when the public service is, for instance, a toll-road simply because the price is flat for any level of demand. But imagine the auction of public land where the buyer of the last plot may indeed be the last one willing to pay such a price-point per acre. Again, the demand in the vicinity of the observed consumption level cannot in all cases be assumed singled-priced. Willingness to pay
is not needed, nor used, for a cost benefit analyses when price and cost are observable and it is plausible to assume that either the marginal cost or the demand are flat. But if it is perceived a high sensitivity of cost or price to the service level, they are needed.

Boiling down all consideration to a single scale is open to criticism on both its normative and methodological implications. On the normative charge, economic individualism is not a precondition for methodological individualism, yet cost-benefit imposes self-interest as the sole driver of individual valuation in order to produce a social aggregate of welfare. Moreover, the choice of a discount rate is a normative choice because it represents the opportunity cost of the stream of net gains, which implicitly is a preference ranking of alternative states of affairs. On that methodological charge, willingness to pay misrepresents individual valuations when the presumed process that elicits valuation–utility maximization–is rendered internally inconsistent in the most common circumstances. Consider for instance, the case of preference-orders susceptible to price changes (as it happens often on eBay or Infomercials) or the case when the total of item-valuations of a basket exceed the budget constraint (evidenced by the sizable outstanding balance in most credit card holders).

Therefore, it is likely that the methodology of policy evaluation will benefit from addressing two weaknesses. First, the commitment to consequentialism seems to unduly constrict evaluation because public action is never exclusively motivated by

---

7 The methodological strategy of studying social phenomena as the aggregation of individual actions does not imply any specific assumption on individual motivation such as self-interest or rational egoism. Recall Weber’s (1922|1978) admonition: “It is a tremendous misunderstanding to think that a ‘individualistic’ method should involve what is in any conceivable sense an individualistic system of value” (p. 18).
the desire to achieve certain ends, but also because the act itself of making policy is congruent with certain principles and duties that we hold dear. Considering this duality of motive, an evaluation of motivations of public action complements the assessment of policy impacts. A subtle form of this duality is observed in political discourse. A coalition or a public official cannot garner political support for a policy proposal sounding half-hearted about it. Political promises are often tempered by the degree of uncertainty born in all predictions of policy impacts; when discourse cannot credibly assure results the uncertainty-gap is filled with a commitment to popular principles. As a result, policy advocates knit political discourse with two rhetorical threads: one alluding to the desirability of the likely outcomes; the second, declaring the principles upheld by the policy itself and the congruence of its likely outcomes with those principles and other core values. The point is that one important reason for policy is the act itself of enacting it; there is a symbolic value in making policy.

There is no need to impose the assumption that all public action is done to advance some democratic principle because modern democracies have formal institutions that regulate the possibilities of public action justifying their character in reference to specific principles. While defining in strict terms what principle means is cumbersome, it is not difficult to recognize that social groups share a few stable precepts—such as the rule of law or free speech and religious freedom—and in the liberal democratic tradition, these precepts are often deemed to be categorical and universal.

Suggesting that policy may be in itself valuable is committing to a deontological theory of evaluation. Philosophers contrast consequentialist and
deontological theories and it is worth clarifying my position in this respect.

Evaluative consequentialism is teleological because according to it the moral value of an act is derived from the value of its purpose, of its *raison d’être*. Depending on the theory of moral value adopted, consequentialism can be utilitarian when hedonism commands valuation, or welfarist when a measure of welfare determines valuation.

In contrast, in non-teleological theories, or deontological, both the act and its consequences are judged by their congruence with a value system. Different deontological theories of moral value will produce different sets of criteria to assess value of policy consequences. Let us stress this point: deontological theories are not “views that characterize the rightness of institutions and acts independently from their consequences”; as Rawls explained, “ethical doctrines worth our attention take consequences into account in judging rightness.” (1971, p. 30). The logic of arguing in favor of a shift from consequentialism to a deontological normative theory of policy evaluation is a logic with which we are all intuitively familiar. Our behavior is often purposive but desired consequences are never our only motivation. We often check the meaning of the act itself against our values and sometimes conclude that restraint is warranted even when the consequences were socially valued. The theoretical challenge is then to characterize such a “value system” or set of criteria well enough to inform policy evaluation.

The second underdeveloped area of policy evaluation is the reconciliation of any normative assessment of policy consequences with democratic values. A common strategy of evaluation is to uncover policy impacts that are at all lights worrisome, even alarming. These normative assessments could be called “front-page headlines” because they are akin to the type investigative journalism that unearths
rotten problems. This strategy comes short for the wide majority of policy problems where problems are more subtle and complex and can only be assessed across multiple dimensions that do not boil down to a single shocking fact. Another common strategy is to use an evaluation method containing a packaged normative theory.

A case in point is cost-benefit analysis that has utilitarianism built-in its method. Let me be more explicit about this, the cost-benefit approach admittedly is a clever contraption; it lets individuals reveal their valuations and indexes them as a price they are willing to pay. By having a common unit the different individual valuations can then be added to a grand social total, just as Bentham envisioned his felicificus calculus. Nonetheless, the normative theory that undergirds the cost-benefit approach is suspect to be at odds with at least one aspect of democratic rule; that of political equality, or more precisely, equal consideration of interests. Specifically, if policy evaluation is to guide policy decisions, the interests of the constituency served by the policy ought to be given equal consideration, but in cost-benefit evaluation individual interests are given asymmetric consideration by construction. Given that: (i) welfare is defined as total surplus equal to consumer plus producer surplus, (ii) that consumer surplus is the summation of individual surplus of all individuals willing and able to purchase the good, and (iii) given that individual surplus is the difference between two prices—the would-pay price and the actual market price--; then, different levels of individual surplus mean that different individuals have different weights as measured by consumer surplus. Furthermore, given that: (iv) the would-pay price is nothing more than the demand schedule, and (v) that the demand schedule is derived from a utility maximization constrained to a
budget; then, for any two equal preference profiles, a larger budget will result in greater willingness to pay. As a consequence, policy decisions guided by cost-benefit analysis seem to accord voice in proportion to purchasing power. What is more, those who cannot afford to pay are not given any consideration, regardless of their preferences or needs. Furthermore, units of producer surplus are the same as units of consumer surplus; implicitly then, cost-benefit analysis grants equal consideration to the producer. Evidently, that too is inconsistent with political equality; particularly, when production is concentrated in a few firms, where each top executive of those firms would be given disproportionately more political consideration than each consumer.

Both stances, the built-in normativity in the method of analysis and the assumed normativity of certain troubling outcomes, hides in the background the normative frame of reference. In the next section I suggest a few conditions for the formulation of such a normative framework.

3. Towards a normative framework for deontological policy evaluation.

The first question in assessing the value of a policy outcome or the policy itself is value for whom? If evaluation is put to the service of democracy, it must provide a democratic answer to that question. While the argument here is has in mind the institutions of representative democracy, they may be reconciled with non-hierarchical forms of participation in policy-making.

From the outset we must clarify that formulating a normative framework for evaluation is not an effort to propose a closed-ended theory of value for policy; this has been precisely the shortfall that we are seeking to correct. Assuming that a
plausible or popular theory of value is universally adequate, as it happens with consequentialism, subverts the notion that democracy can accommodate a plurality of values, and in the specific case of consequentialism we have argued that it misses entirely the duality of motive in policy making.

A first approximation could be the formulation of a set of minimum standards that policy outcomes and process should satisfy if they are to be considered legitimate. This is precisely the strategy of Public Value Failure analysis (Bozeman, 2002; 2007; Bozeman and Sarewitz, 2005; Fenney and Bozeman, 2008; and the case studies discussed in Bozeman and Sarewitz, 2011).

After demarcating the outer terrain in which a policy is situated, a second approximation to the normative frameworks that we seek to sketch could be in narrowing the preferable set of values and map them onto the practical aims of policy. Such an approach was proposed by Frank Fischer as a “critical policy evaluation” (1988, 2003). His project is certainly coterminous to the deontological evaluation proposed here; however, Fischer sought to establish such a mapping distilling social values from the political philosophy of the Frankfurt school. In the actual practice of policy making this choice of a set of social values will meet resistance from those who prefer a different set of values, particular those consistent with classical liberalism. For instance, the values from the political philosophy of Rawls (1971, 2005|1990) as condensed in the “difference principle” and “overlapping consensus” may lead to an evaluation of policy in regards to “primary goods”—resources and opportunity that citizens deploy and use in pursuing their conception of the good life. Also within the liberal tradition the mapping could be done against the enunciated capabilities of Nussbaum’s “capabilities approach” that
refined Rawlsian (Nussbaum, 2000). A possible weak link in Fischer’s approach lies in that mapping a well-defined set of values onto policy goals may inadvertently validate some of the justifications of such goals. For instance, assessing developmental policies (that rely heavily on innovation) as a vehicle to alleviate poverty using Nussbaum’s approach—which incidentally is done by official UN assessment of the Millennium Development Goals—implicitly validates the view that poverty is something that can be solved with innovation. Those people called poor may instead see poverty as derived from their historical political marginalization, but those values are invisible to developmental policies center on mosquito webs and water filters.

Consequently, a third approximation to a normative framework for deontological evaluation ought to take seriously the values of the publics impacted by policy. The legitimate representation and aggregation of those values is precisely the historic challenge that various forms of representative democracy have tried to resolve. This means that two critical issues are at stake, how to understand representation, and how government derives legitimacy for its authority.

To respond to the first question I draw directly from the discussion of representation in Mark Brown’s *Science and Democracy* (2009). Brown is critical of “theories of correspondence” that have characterized representation. Not only political but also scientific representation too, commonly thought to be different kinds of representation, have actually been conflated by the liberal democratic project. When no specific mechanism (like voting) can project the image of the “will of the people” to scale, as an architect’s design, neither view of government as a delegate or as a trustee are adequate—this is a parallel argument to the seminal work
of Pitkin on representation (1967). Brown conceptualizes representation as a
dynamic process that results from the functioning of a whole ecology of institutions
of representation, each of which satisfies one or more of several critical functions,
namely: “authorization, accountability, participation, deliberation, and resemblance”
(Brown, 2009, p. 206). The thickness and heterogeneity of this ecological space,
extending the natural analogy, determines the resilience of democratic representation.

To respond to the question of government legitimacy, Thomas Catlaw who
in Fabricating the People (2007) offers an original perspective studying the legitimacy
question with respect to the bureaucracy. Like Brown, Catlaw rejects the theories of
correspondence, but locates the culprit for the misconception in the ontology of the
sovereign of democracy, the people. The liberal democratic project has been
ultimately a project of creating a political identity in the sense that representation of
the sovereign as one person, one will, and succeeds only insofar as the sovereign is
believed to exist and is given its coercive authority by every citizen. Moreover, this is
only possible if the political self-concept of every citizen can accommodate such an
entity. The inevitable consequence of the construction of this political national
identity is the demarcation of its contours; that is, the construction of its negative
which is the act of political exclusion. Exclusion is thus as fundamentally democratic
as is representation. The contemporary crisis of administrative legitimacy is brought
about by the fracture of the political identity. Once the national identity is fractured,
exclusion is equivocal, almost capricious, and thus it slips into further fractious
conflict. The current legitimacy crisis may then explained, at least in part, by the
social tensions that became apparent since the 1960s (one could imagine that a
similar and perhaps deeper fracture led to the Civil War). I would hazard to add that
it has been also co-produced by the rise of information technologies (c.f. Castells, 1996). Catlaw concludes that government and the bureaucracy will only undermine further their legitimacy if they insist in fabricating that “unum” of e pluribus unum. Government and policymakers that admit the defeat of the national political identity project must anchor their legitimacy in a multiplicity of institutions that have retain their legitimacy independently, or in Brown’s terms, must anchor policy in the ecology of democratic institutions of representation.

The political philosophy of Brown keeps the term representation but redefines it, Catlaw prefers to vanish it, but both rely on a robust and heterogeneous ecology of institutions to preserve democratic life. I submit then, that deontological evaluation would sharpen the examination of values relevant to policy by mapping them onto the institutions of representation and characterizing their dynamics. This is indeed similar to the policy analysis method that Bozeman named Public Value Mapping (2003) with the caveat that, to the light of the foregoing considerations, the characterization of values must be careful to discern values that emanate from theories of correspondence in representation and those that re-fabricate the people as a national political identity. For instance, deontological evaluation would distrust deriving values form polls and surveys given that they systematically boil down heterogeneous identities to a single menu of responses.

4. Conclusion.

---

8 We will not answer here the question of what are these institutions and how can we tell which are to be nurtured and which to be stunned. We might suggest that the design of those institutions may benefit from the work on the stabilization of social boundaries, called “boundary work” (Gieryn, 1990, Jasanoff, 1990) and “boundary organizations” (Guston, 2000, 2001).
The motivation for a deontological evaluation should not be confused with the now popular characterizing of politics as local and its derivative prescription for evaluation to be sensitive to local and contextual needs, although I see those sensibilities as necessary of deontological evaluation. The motivation lies instead in the need to reconsider the political power of the various stakeholders in a policy, particularly those groups that privileging consequentialism excludes from deliberation. Deontological evaluation must elevate the voice of traditionally marginalized technical and political representations. What is original of this proposition, the mapping of values into the various institutions of representation, could benefit from the methods of institutional analysis that have thus far been developed into a robust toolkit of analysis (Ostrom, 2005).

This project starts with a critique of the epistemological and ethical gaps that have grown within the discipline of evaluation, and I would add, gaps that have grown in despite of the increasing sophistication of its methods and techniques. Nonetheless, I should stress that, like Fischer’s “critical policy evaluation” (1988), and Schneider and Ingram’s (1997) “policy design,” deontological evaluation does not discard the conventional practices of evaluation; rather, it seeks to complement them, and ultimately, it seeks to assist other democratic institutions in restoring legitimacy and trust in government.