Accountability for Creativity and Discretion

© 1990

Michael O'Hare
Mark H. Moore
John F. Kennedy School of Government
Harvard University

*DRAFT: Please check
with the author before citing this paper.*

Introduction

We may take as moribund, if not mythical, the Weberian manager who adds to politically legitimate instructions from above only the expertise inherent in a technical specialization of some sort. As an object of study he is boring and unenlightening, and as a reality he is impossible on grounds of information handling capacity and productivity: no supervisor can know enough to control subordinate managers in this way, and no organization can afford to sacrifice the creativity and imagination of its staff.

In his place, we find a more or less effective manager who has political as well as technical or administrative concerns; who is proactive and experimental rather than reactive; who explains the situation to her superiors and may even present finished product for approval rather than seek permission before starting; who seeks and fashions political mandates rather than waiting by her in-box for instructions.

This manager has been called a "Theory S" [for *strategic*] or a political manager, usually approvingly, in scholarly discussion, and a leader and effective executive in popular media. Occasionally a William Casey reminds us that the realistic limitations on supervision and control, to say nothing of explicit encouragement of independence for public managers, create real risks of abuse. Active, creative, strategic management at any level raises important and persistent questions usually treated under a heading of "accountability." Indeed, what Casey wanted to create was an umbrella under which he would not be accountable to anyone.

The ants in T.H. White's *Sword in the Stone* were guided by the twin principles, *Everything not forbidden is obligatory* and *Everything not required is forbidden*, but public officials who are supposed to do their own thinking have much more discretion. This discretion raises a continuing challenge of designing effective, appropriate, and humane systems of accountability. Indeed, an accountability structure remains the most important lacuna in the Theory S managerial model. Where the theoretical bureaucratic model could at least hold up the ant formula as an ideal, suggesting that accountability problems for managers might be dealt with by moving towards ever-more-specific anticipation of options and instructions, it seems to point in exactly the wrong direction for the strategic manager. If good management brings imagination, creativity,

and independence to public service, it doesn't seem promising to pursue it by trying to outguess and constrain the manager in advance.

It must be noted that detailed instructions and constraints did not in any case constitute a workable accountability system under a traditional bureaucratic regime. If subordinates are capable of independent action, rules and standards require observation of their behavior in order to impose sanctions for error, and bureaucratic organization does not greatly simplify this task. White assumed it away for his ants by giving them instant telepathic communication with headquarters through their antennae. The accountability problem of strategic management is not that managers used to be accountable and now are not, it is that philosophical and metaphoric problems involving the idea of responsibility and guidance have been added to practical problems of observability and auditing.

Accountability is itself a complicated concept, embracing not only languages to describe managerial behavior and standards for performance, but also the system of discipline and rewards for good and bad performance. In its most rudimentary form, it suggests a system of "conservation of blame," whereby anything that goes wrong can be traced to someone in order to demonstrate personal culpability. In more useful form, it evokes a complex and subtle network of assessments, communications, and signals that direct the attention and work of individuals toward appropriate ends, including ends of risk-taking and experimentation. The better a system of accountability is, the more it fits and overlaps a general system of management.

The standard analysis of managerial accountability begins with the principal-agent model, involving two parties, P and A, with distinct utility structures, who have agreed to act, for whatever reason, so as to advance the interests of P. P might be A's boss, or A might have been elected by voters P. Agency models describe schemes by which A's interests might be aligned with P's, such as raises for acting as P wishes and punishments for pursuing his own goals at a cost to P. Elaborations of the model examine the problems of observability and uncertainty -- how can P know what A is up to, or second guess the expertise for which A was presumably hired in the first place, in order to impose the optimal rewards and punishments? what if neither P nor A can be sure what actions will really advance P's interests?

The behavior organized by an accountability relationship is, in similarly abstract terms, a three-step process. Step one is the formation

of a contract. Step two is the observation of some indicator of A's performance by B, and step three is a change in A's circumstances (typically some sort of reward or punishment) whose relationship to the observation is given by the contract. Step one might occur only once, and the second and third repeatedly; alternatively one might observe continuous recontracting through the discourse of observation and response, or explicitly.

Although any utility structure can be imagined for A in a model of this type, including the possibility that the agent wants nothing beyond the principal's welfare, it lends itself, and implies from initial conception, a situation in which A is disposed to steal from P, or wishes to misuse assets or discretion in the service of values different (though not necessarily completely different) from P's, or is just inept. The agency model strongly suggests, or assumes, a defect-finding or error-correcting conception of supervision, which in turn assumes the existence of an *a priori* specification of A's correct behavior. Defect-finding is affectively inappropriate in general, but if agents are supposed to have ideas and be creative, the mismatch between managing on failure to meet a known standard and the intention of motivating repeated excursions beyond the standard is important.

Furthermore, while Casey or an embezzler are the dramatic symbols of agency failure in government, missed opportunities to excel and well-meaning ineptness and ignorance are probably much more common and more damaging. While an accountability structure must assure the public protection from intentional and cynical misuse of power, emphasizing purposeful offenses over unintentional failures to excel carries the implicit message that public officials are prone to steal and subvert, and may even stimulate a sort of sheep-for-a-lamb response.

To advance from "catching bad guys and punishing them," beyond even "catching good guys and giving them raises," to "improving the performance of government," requires that we enlarge the model of accountability beyond error-finding. What purposes would an accountability structure serve in a world of mostly well-meaning and competent public officials?

The question people most ask about their work was immortalized by Ed Koch: "how'm I doin'?" The next is "how does my boss think I'm doin'?" Not far behind these is, or should be, "what can I do about that?" Our primary instinct is for evaluation, but even modest sophistication and self-awareness teach us to seek guidance and advice. Rationally, the first question is useless, the second is at least predictive,

and the third can really do something to help create value. Affectively, the order of importance is probably the opposite, and there may be a moral right to have the first question answered even if it can't be shown to lead to productivity gains.

The agent, then needs to have some account of how his performance compares with what he might be doing, and he needs it both on Maslovian grounds and as a way to know what to do next -- that is, to know how to allocate his time and attention.

The principal needs similar information, usually about several agents at once. Here the accounting has less affective value, but still serves allocation processes, some concerned with assignment of people to tasks and others concerned with allocating training, capital equipment, and personnel to tasks and opportunities. And since the principal is concerned with everything that might affect the agent's ability to accomplish work, he, too, wants the agent to have the accounting suggested in the foregoing paragraph.

Accountability Space

The system of accountability in any real organization can be characterized on eight critical dimensions, most specifying the information on which decisions will turn, the others the decisions that may be made. An particular accountability structure thus implies a value on each of these dimensions. A single organizations inevitably have several such systems, different from each other on several dimensions, to accomplish different purposes, and real accountability structures may be found over a surprisingly wide volume of the implied eight-dimensional space.

Objects

At the outset, one cannot give an account generally; a manager is accountable *to someone*, and therefore not to others. Who is entitled to demand an account of an agent's behavior determines, to a large extent, who the principal is with regard to the dimensions at issue. It's easy to forget that while subordinates are generally accountable to superiors for all sorts of things, there are other accountability relationships in most jobs, and the boss may even be specifically disempowered to demand

certain kinds of accounts. Professionals, to offer the most familiar example, are accountable for process (see below) to the governing bodies of their professional associations; a prosecutor is similarly accountable to the District Attorney for one group of decisions, ~~and~~ to primary quality ~~of any agent's accountability is the person or persons to whom he is accountable. Only certain parties may demand an accounting~~

an appellate judge for others, and to the voters for still others.

Stages of Production

According to a familiar and useful model, organizations (including government agencies) use *inputs* (factors of production like labor and paper clips) in *processes* whereby value is added and the inputs transformed, to generate *outputs*: the department of public works deploys drivers and trucks (inputs) on routes and schedules (process) to get some number of tons of trash from curb to landfill each week (output). There is no intrinsic value in hauling trash around, however; we only do it because it leads to better public health, a better environment, reduced use of plastic through recycling, and similar *outcomes* that are usually (i) remote in time and space from the organization at hand, thus difficult to observe, (ii) due to the actions of many organizations, thus difficult to link to any single organization's decisions, and (iii) more obviously tied to basic social values, thus more central to our interest.

Each of these four stages permits observation for purposes of accountability, though inputs and outputs are frequently easier to measure quickly. In the private sector, both usually leave a trail of accounting information through market transactions; public outputs are less commonly bought and sold discretely but many are subject to weighing, counting, etc. Processes are usually more difficult to observe and contrast, and outcomes, especially public outcomes, have the qualities of joint cause and distance from the fact mentioned above. These contrasts set up a constant underlying tension between observing cheaply and unambiguously, and observing what we care about.

Census and Sample

Whatever stage of production is observed, more or less of it may be noted and judged. In special cases, we try to check everything: my bank is quick to inform me that ATM deposits are counted by two employees at once, each serving as principal and agent for the other. Accounts are audited, checking each discrete transaction. But backstopping

of this kind is conceptually without end: as Zero Mostel used to ask in a political song of the late 40's, "Who's gonna investigate the man who investigates the man who investigates me?" Accordingly, principals need ways to control the amount of information the accountability system has to process.

The first means of doing so is to sample. In addition to being less costly, sampling is much more interesting, complicated, and instructive than exhaustive review. Samples may be taken at random, or data that is especially informative may be sought out: the commissioner of public works might think trash collection on rainy days, when the system is under stress, will be especially illuminating of the general performance of his crews. If the sampling rules can be inferred, a game is immediately created between principals and agents in which the principal's inference can be systematically subverted.

At the same time, the sampling rules constitute a channel by which the principal can signal the agent what aspects of his performance are especially important. And a selection may be made not only on logical but also on symbolic or metaphoric grounds: in the context of consequences (see below), an especially poignant or memorable success might be chosen for acclamation and a big reward, rather than distributing bonus money thinly across many claimants.

A hybrid system is frequently set up in which the transmission of information from agent to principal is suspended in midair, so to speak. Agents may be obliged to accumulate exhaustive records (a census) that no-one will ever read unless suspicion of exceptional performance (usually underachievement, unfortunately) triggers a detailed and selective review, or merely to support random exploration for statistical samples. Especially as computers make it relatively less expensive to accumulate records (of at least some kinds) and to find them for inspection, a scheme of this kind is attractive at least on theoretical grounds. While we can't review all actions, we can in concept review any of them, and if punishments and rewards can be increased to keep the expected value of being caught at the appropriate level it should be possible to manage much accountability this way.

Unfortunately, sampling with atom-bomb sanctions is not so easy in practice. In the first place, it's not usually possible to impose appropriate punishments; no matter how salutary one might think the spectacle of courtmartialling and discharging the private with a dirty rifle would be for the observing troops facing a one-in-a-thousand chance of having their own weapons examined, the public will not stand for the in-

dividual injustices inherent in such a scheme. In the second place, the demoralizing effect of such official random sanctions -- possibly similar for apparently disproportionate benefits; should the soldier caught at random with a perfect target score be promoted to captain? -- are likely to outweigh the desired effects.

Aggregation and Detail

The other conventional means of limiting the principal's attention to the agent's behavior is by aggregation: the rules of financial accounting can be thought of as a system of aggregation rules, specifying which numbers may be added together and reported as a single fact (the amounts owed to each of a thousand bondholders), and which must be kept separate (debt from equity).

As I will argue below, aggregation is a critical factor in effective supervision and accountability. It is itself multidimensional; performance measures may be aggregated over time, suppressing quarterly reports in favor of annual aggregates; across people, judging the performance of teams rather than individuals; across processes, as usually happens when outcomes are observed; and in still other ways. Renaissance explorers and diplomats were told to "serve the interests of the crown" and sent off to work for months, even years at a time; modern ambassadors check in by cable and telephone before every meeting and fly home for instructions at the drop of a presidential poll standing.

Unfortunately, a characteristic tension draws us toward aggregation for purposes of consequences like firing people or electing the President (only a fool fires employees after the first mistake), and toward disaggregation for purposes of improving performance. McDonalds, for example, finds it useful to observe forty dimensions of each store's performance (litter in the parking lot -- 1 point; quality of soft drink, 5 points; quality of Egg McMuffin, 12 points; greeting, thank you, etc., 3 points; etc.) four times per month, and report the detailed results to the store manager despite having the aggregate profit information that might seem to summarize outcomes in that context. We say *unfortunately* because for reasons unique to government, the existence of detailed observations can make it difficult to link consequences to aggregates. We will return to this tension below.

Explication and Implication

Locations on the foregoing dimensions determine what people will be accountable for. But how do they know? The terms of accountability may be fixed in any of three ways, with further variations possible. Some instructions and standards are always explicit, in the form of job titles, statutory language, written contracts, or even conversations between principal and agent. Others, again always, are implicit: political appointees are expected to help re-elect their appointing officer, but they don't usually need to be told to do so. All public officials are presumably supposed to report wrongdoing and discourage it, and to care for their subordinates.

Some further instructions are not only implicit but tacit. Not only does no-one formally tell you to try to find a job for the nephew of the speaker of the house, but they will deny, if asked, that you are supposed to do so.

In general, punishing consequences are attached to violations of explicit conditions, while promotions and rewards tend to follow achievement of implicit goals. This may be because punishments can be challenged by litigation, and litigation favors an explicit record. The convention leads to occasionally absurd results, as in the recent case of the National City, California policemen who were not disciplined for cheating on a promotion exam because they had not been told not to cheat.

Post and Ante

A critical distinction in the context of the Theory S manager is that between accountability before and after a managerial action. Some bosses like to be shown finished product, including mistakes. Others like to be asked in advance. Pathological cases exist, like the principal who won't commit either way before a risky enterprise, but savages the authors of failure.

The difference is important when the agent can be expected to "grope along" in a new enterprise, and cannot therefore give a specific account of what he will really do; of course, this groping along and experimentation is of the essence of Theory S management. A creative manager might get *ex ante* guidance in outcome, value, or aggregated terms, and may also benefit from teaching and advice that have only a

distant relationship to terms of accountability, but he probably cannot operate without an understanding that much of his contract will have to be specified retrospectively.

Inclusion and Exclusion

The discourse of accountability falls into general categories of obligations and prohibitions. A principal and agent may agree on what the latter should do, or may emphasize what he should avoid. The way sins (and successes) of omission and commission are allocated among explicit, implicit, and tacit specification will have much to do with the likelihood that an agent will charge forward or just keep his head down.

Consequences

Finally, critical to any accountability system are the consequences that agents may encounter following the accounting itself. One important categorization of these is between punishment and reward; a pay cut is significantly different from foregoing a bonus of the same size. Another is between economic and other consequences, a distinction often made confusing when economic rewards matter more as signals of merit than they do *per se*.

Public principals often feel that the range of consequences they can attach to performance is too limited; civil servants are hard to fire, raises are constrained, promotions too slow, and a bigger office too hard to wangle out of GSA. Worse, some genuinely perverse incentives frequently apply: the reward for cutting your budget this year may be that you have to live with it indefinitely, a treacherous response in a world where managing more people is a good proxy for prestige and highly correlated with pay and promotion.

Much unnecessary constraint results from an uncritical bunching of different kinds of consequences. Associate professors are commonly rewarded with an indivisible package that includes tenure, a larger office, a raise, more secretarial help, and more authority (over apppointments, for example). Excellent schoolteachers are rewarded with a promotion to assistant principal -- that cannot be separated from management authority for which they may not be qualified. In contrast, the

Army has established a promotion track (for "specialists") that allows higher pay and other perquisites to be awarded without also awarding command prestige.

Third Parties, the Fishbowl, and the Bottom Line

The dyadic agency model involving a principal and her agent has served to this point as an adequate framework, even when the principal is a group or fractionated, and the agents legion. However, public managers' accountability is of interest to a distinctive set of third parties that cannot be properly modeled either as more principals or more agents. These are the press and members of the other party. What is important about both is that their interests are aligned with the dyad's only at the highest and most abstract level -- Republicans, Democrats, and Mike Wallace all favor the flag, motherhood, apple pie, and a great future for America, but Republicans are supposed to try to get Democrats out of office, and Mike is supposed to afflict the comfortable of both types. Indeed, if the Democrats adopt a Republican position, moving toward the center, the Republicans are supposed find new grounds on which to contest them, and if the subject of Mike's exposé reforms, he is supposed to go after someone else.

Many government agencies have antagonists independent of the media and of party: legal aid offices look to many landlords like the enemy, and advocates of solar power might find it advantageous to make the state energy office look inept or prejudiced.

Private sector organizations have third parties to their agency arrangements as well: the McDonalds manager and her employees have to look over their shoulders at Wendy. The critical difference is that with specific and narrow exceptions, McDonalds' internal accounts are private while government lives in a fishbowl, illuminated by sunshine laws and at constant risk of a swoop by a reporter wielding the freedom of information act net. The information in government's accountability system is available to a variety of people who share the interests of neither the principal nor the agent -- who may be actively, and appropriately, antagonistic. While critics may not have the power to fire or cut pay, they often control or influence public repute, and public repute is a relatively important motivator for many government officials.

Adding to the power of third party criticism is the close relationship between public approbation and what democratic government is

supposed to do, which is to attract public votes. The businessman criticized in the pages of *Forbes* may wince, and his stock price may even register a bobble, but increased profits are a powerful and legitimate rejoinder. The public manager has no such independent measure of merit to point to: if the public thinks poorly of him, he is *prima facie* a poor public official.

Consequently, a public system of accountability is subject not only to the perverse incentives and uncertainties of all such systems, but also liable to exploitation by parties trying to injure the organization itself. This would be bad enough, but such exploitation, combined with an extremely robust and not unreasonable convention about what news is and what it isn't, has a more distressing effect. The convention is that news is new, finite, unusual, and exciting. Criticism of party A's oractice by party B is conflict, which is always exciting. The other rules favor discrete events rather than long-term performance, and in a world where competence and success is typical, errors and failures.

The result is an overwhelming selection bias whereby the portrait of government behavior created by the media who, willy nilly, portray it to the public, is a tangle of failures, crimes, and criticism. On days when no scandal has been turned up by reporters themselves, opposition party representatives are likely to be captured viewing with alarm and criticizing. It's not surprising in consequence that the public should demand a system of accountability whose primary task is to assure exposure of what looks like wholesale wrongdoing. Unfortunately, a system of accountability *primarily* oriented to catching and punishing malfeasance is at odds in several important ways with the goal of advancing average performance beyond a minimum standard. And the underlying implicit assumption -- that government is a nest of thieves and nitwits -- is in conflict with the idea of Theory S management itself.

Aggregation and Example

Recall that we have identified at least four distinct functions of an accountability system. First, it should make agents' own actions and their consequences visible to the agents themselves, and do so in such a way that they can revise and correct their own behavior. Second, it should make the behavior of agents visible to principals in a way that supports (i) allocation of guidance, rewards, and punishments, and (ii) efficient deployment of resources, such as reassigning people who aren't doing well at their assignments and providing resources to those who can best use them. Third, it should make possible comparisons between competing officials, especially incumbents and candidates, that lead to informed public decisions like elections. Fourth, it must assure a reasonable citizen that public officials are generally doing what they are supposed to do.

Complicating the demands on any scheme is the familiar fact that "everyone is a middle manager," the agent of principals and the principal of agents. To the extent that agents and principals need different accounts in a single relationship, this means that any individual in a properly managed organization will be receiving different kinds of information at the same time to support different roles.

These functions are characteristically served by observations and consequences from widely different places in the accountability space described in the earlier section. For example, agents need detailed, disaggregated information about what they are doing and what effect it is having. To give this information weight, it needs to be associated with differential consequences, but the range of rewards and punishments associated with a unit of feedback -- evaluation of a week's project, for example -- must be limited to a range consistent with the agent remaining in his current position and maintaining confidence and optimism. Furthermore, even large and costly errors must often be treated on the basis of the intentions of the agent, and in light of what good general practice seemed to dictate at the time, or the agent will simply stop displaying the imagination and courage the organization requires.

Principals need much more aggregated information about the same events, and often have reason to collect it, and be seen to be collecting it, about teams and groups rather than individuals. The aggregation has two purposes. First, it limits the information that must be processed to manageable amounts; second, it assures that errors are handled together with the successes and learning for which they are essential.

Principals are also more concerned with intrinsic qualities of agents than are the agents themselves. A taciturn and pessimistic person can't do much about his gloomy nature, and won't be helped much by being reminded of it, but his boss needs to know that he shouldn't be assigned to train the summer internes.

The principal must also be concerned with the signals to agents implicit in the information gathering process itself. It's a familiar maxim, for example, that simply collecting information about a dimension of performance and being seen to do so (e.g., by returning it to the agent) will assure more of whatever is being measured (or less, if the dimension is negative).

The language in which performance is best described also varies with the purpose and recipient of the report. An agent may want to describe his behavior to a principal in a symbolic example rather than an exhaustive, dispassionate table of data, especially when he is trying to get his terms of accountability revised in light of some new discovery about the agency's work. This kind of symbolic speech is especially common when the agent is a candidate or senior official and the principal is the public, and is probably essential to a proper account of government work.

These different forms and terms of accountability are each indispensable. Should a program that gets a thousand mothers off welfare, in which one district office manager gave two no-show jobs to his cousins, be evaluated on the basis of a video history of one of those women, the aggregate statistics of the process, or the abuses? Obviously, the answer is, "all of the above." But there are three different evaluations for three different purposes involved. Unfortunately, it is quite possible that the wrong observations will be used for each. The two no-shows, in response to which the program director could refine his own practice, could well become the public image of the whole program, to no good effect. The statistics that would enable the welfare commissioner to compare programs with similar purposes might be used only to allocate raises and promotions, and the case history might become an internal mascot, reassuring to be sure, but leading to no practical evolution or learning.

The challenge to the Theory S manager, then, is not to find an accountability system for his agency, but to establish several systems and match their information demands and collection to their appropriate purposes. One underlying pattern of these systems is likely to be that decisions that allocate more assets and resources should be based on more

aggregated information than "smaller" decisions. Thus, hiring and firing should reflect knowledge of a long history of an employee's performance at many different tasks, while a larger office or a new computer could result from more detailed observations. Whether a program should be maintained should depend on an account that involves the whole agency's performance over time, while incremental budget adjustments might weight recent performance at particularly salient tasks more heavily.

These distinct systems need to be differentiated not only on information packaging but also in terms of consequences. In a sense, each of them constitutes an internal market with its own medium of exchange. Salary, position, capital equipment (office, computer, furniture, etc.), and more than one kind of acclaim can be allocated independently. Separating these "coins" enriches the system of accountability, allowing assurance that good and bad behavior will be attended to in a meaningful way at different levels.

A second general principal is the separation of accountability for localized failures to discharge general duties from accountability for broad failure to meet specific missions. In the example above, nepotism -- even theft -- are failures to do what any public official should do, and not (however serious they may be) at the strategic center of a welfare/training agency's mission. (They might be at the center of an Inspector General's office mission.) One can imagine a whole agency being crippled for months or years by a "scandal" of this kind, even though if we knew in advance that its mission simply could not be accomplished without a few abuses, just as coal cannot be dug out of the ground without some loss of life, we would still think it worth having. The recent history of the Job Corps, whose reputation and capacity to function was severely damaged by criticism of its contracting procedures, exemplifies the costs of failing to keep accountabilities in their proper categories.

This separation, which usually takes the form of protecting of an agency generally from attacks that properly go to a localized or strategically peripheral problem, usually involves a manager at least one level above the director of the agency at risk. When the hue and cry is up and your agency is publicly embroiled in a delicious scandal, survival probably depends on your boss -- not you -- insisting that the affair be kept in perspective, and explaining why. But it also depends on having established a clear picture of the agency's performance at its central mission before the storm breaks, and this requires that the agency manager take

an active role in creating that picture and diffusing it. Outside parties will look for behavior -- and misbehavior -- that "anyone can understand," which too often is the sort of abuse for private gain that any agency might harbor, and too rarely is the specific sort of public value the agency itself was mandated to create.

The third principal is to ensure that each of the several systems can accomodate, and respond to, evidence of exceptional success as well as failure. This criterion is so rarely satisfied that it's hard to find examples. For example, all rules and standards in prohibitory form fail the test. Performance evaluation contracts, that usually specify tasks to accomplish and deadlines, allow one to come in ahead of time and under budget, but rarely characterize what it would mean to exceed expectations at the task itself. For an agency as a whole, the contract with a legislature and the voters that anticipates overfulfilling their expectations is an almost entirely new mode of discourse that awaits imaginative managers' attempts.

This is precisely when "strategy" — with an associated set of performance measures fits in.