Center for Technology Innovation at BROOKINGS

June 2015

The politics of federal R&D: A punctuated equilibrium analysis

By Walter D. Valdivia and Benjamin Y. Clark

Walter D. Valdivia is a fellow in the Center for Technology Innovation at the Brookings Institution.

Benjamin Y. Clark is Associate Professor at the Levin College of Urban Affairs, Cleveland State University, and Executive Director of the Great Lakes Environmental Finance Center.

EXECUTIVE SUMMARY

The fiscal budget has become a casualty of political polarization and even functions that had enjoyed bipartisan support, like research and development (R&D), are becoming objects of controversy. As a result, federal R&D is likely to grow pegged to inflation or worse, decline. With the size of the pie fixed or shrinking, requests for R&D funding increases will trigger an inter-agency zero-sum game that will play out as pointless comparisons of agencies' merit, or worse, as a contest to attract the favor of Congress or the White House. This insidious politics will be made even more so by the growing tendency of equating public accountability with the measurement of performance. Political polarization, tight budgets, and pressure for quantifiable results threaten to undermine the sustainability of public R&D. The situation begs the question: What can federal agencies do to deal with the changing politics of federal R&D?

We turn to history for guidance. We apply punctuated equilibrium theory to examine the last four decades of federal R&D, both at the aggregate and the agency level. The series studied are characterized by a general stable upward trend that is primarily due to the decentralization of the system; each agency requests an R&D level based on its own funding history. We also observed budget leaps or punctuations but they do not appear to have lasting effects: budgets return, in time, to their trend. These historical lessons suggest that agencies are better off securing stable growth in their budgets in the long run rather than pushing for budget leaps. As the politics of R&D are stirred up, agencies are sure to find that proposing punctuations is becoming more costly and risky.

Agencies with small R&D units have kept these units inconspicuous in their total budget requests while research mission agencies have self-promoted and publicized any positive impacts in order to garner public support for their R&D budgets. With tighter budgets, agencies will have to double down on these strategies but will be well-advised to avoid inter-agency contests for funds that could degenerate into contests for political influence. In this environment, calls for doubling any agency's R&D budget in a short period of time, however well-intentioned, are misguided. What is

more, agencies asking for budget increases above the norm will invite growing demands for public accountability from the executive and legislative branches.

This accountability pressure is likely to be operationalized in the language of returns on investment and more generally, the formal quantification of causal links from public R&D to organizational outputs and to societal outcomes. Because research, with few exceptions, cannot be linked in neat causal chains to outcomes, this approach to evaluation is more adequate for development programs than research ones and agencies should resist it accordingly. Rather, research should be evaluated by the integrity of the scientific process and by how sensitive it is to public values and social expectations.

In the past, the stability of R&D spending resulted from the character of the budget politics. In the future, stability will need the stewardship of R&D champions who work to institutionalize gradualism, this time, in spite of the politics.

1. INTRODUCTION: HOW TO ASSIGN PRIORITIES?

Political polarization, tighter budgets, and emerging conceptions of public accountability, are forming a perfect storm that threatens the future stability of public funding of research and development (R&D). Large-scale federal support of R&D was unprecedented before World War II; since 1953 the R&D budget has grown at the relatively fast pace of 3.74 percent a year. While still increasing nominally on an annual basis, its growth rate has fallen to near zero once discounted for inflation. Policymakers from Congress, the executive, Office of Management and Budget, and federal agencies, as well as civil society's advocates of innovation, must confront this reality and plan strategically how to sustain a base-level of public support for R&D. We present here an analysis of the historical budget data to understand the nature of R&D's gradual growth as well as the occasional bursts of additional funding. This analysis reveals the politics of R&D budgets and informs the role federal agencies, universities, and laboratories can play towards maintaining a stable if modest pace of growth of public R&D so that society can enjoy the benefits derived from this public spending.

Public R&D has historically received bi-partisan support. Each Congress has a few stewards who propose to accelerate this kind of spending. As recent as March 2015, Senator Dick Durbin (D-IL) proposed a steady increase of five percent per year (above inflation) for 10 years in five agencies: the National Science Foundation (NSF), and the science units of the Departments of Energy (DoE), Defense (DoD), National Air and Space Administration (NASA), and the National Institute of Standards and Technology (NIST). This announcement followed a similar funding increase proposal made by Durbin a year ago for the National Institutes of Health (NIH). At a five percent growth rate, it would take 14 years to double the current funding level. These proposals are reminiscent of and probably inspired by the successful lobbying of Congress to double the funding for the NIH between 1999 and 2003.

In effect, federal R&D has a well-organized lobby. Comprised by university associations, notably the American Association of Universities (a group of 62 research universities) and the Association of Public and Land-grant Universities (representing 209 public universities), as well as scientific associations of which the American Association for the Advancement of Science, publishers of the journal *Science*, is the most prominent, the R&D lobby is very active exhorting policymakers and informing the public about the vital importance of R&D to national prosperity.

Supporters in Congress and well-intended interest groups must confront the hyper-partisanship politics prevalent in Washington D.C. and the fiscal crisis it has engendered. Exemplar of this crisis is the stand-off between the executive and Congress that led to passing the Budget Control Act of 2011 with a misguided provision for a *budget sequestration* that designates a specific amount of cuts for each department instead of capping the total fiscal budget. Consequently, the government cannot reallocate resources on the basis of urgency, impact, or any other sensible criteria. No department and no function have been spared, not even the traditionally protected defense budget, from passing through the chopping block. Polarization also renders sequestration unlikely to be replaced with a more sensible measure in the near future. Rather, it is fully expected that piecemeal solutions will be introduced on each annual budget resolution—such as the Murray-Ryan 2013 budget deal (Lawrence, 2015)—granting the government only partial latitude to allocate funding according to its priorities but hardly allowing an overall increase in federal spending.

It is thus realistic to expect that the future of federal R&D will be somewhere between a best-case scenario in which it grows pegged to inflation (keeping its purchasing power) and a worst case scenario in which it declines. This situation is changing the politics of the R&D budget, because when the size of the pie is fixed, any agency seeking a bigger slice triggers a zero-sum game with the other agencies and invites calls for greater accountability from lawmakers. The new budget politics beg the question: How should the federal government assign R&D priorities?

Economists of innovation have taken on a financier's approach to answer this question: Find the return on investment (ROI) of every R&D dollar. This approach imagines the federal government as a shrewd mutual fund manager building an investment portfolio and allocating moneys to maximize the return for a given risk level, or minimize the risk for a given expected return. This has been the dominant approach to study the federal R&D, not only for its conceptual clarity and the resemblance to the practical world of investment finance, but also because policymakers understand this language and use it to demand accountability from agencies and R&D contractors. They customarily demand to know the ROI at the organizational level (e.g. per university), sometimes at program level (e.g. per each of the NIH institutes or even per NSF directorate), and in the most audacious of cases, at the grant level—in fact, some legislators have suggested that each grant should be certified to have a positive impact on the economy.

We believe that estimating ROI is a useful approach to understand R&D priorities when the government decisions follow a comprehensive rational approach, when the public interest can be clearly discerned from the available choices, and when its dictates are unchallenged and carried out faithfully. Advocates of this approach may argue that they are aware of the vicissitudes of any good policy idea offered to any good working democracy, but that their solution to this problem could set a normative framework against which actual policy can be evaluated. This would be feasible if only scholars were the kind of people who quickly agree with each other on normative frameworks. But any estimation of returns on research investments at the organization or program level is sure to rely on many assumptions—some of which will be as questionable as assuming the government works like a mutual fund—and will have to compete with just as many credible alternative estimates. In fact, the real world of investment finance confronts this very problem, for there are as many opinions on how to build an investment portfolio as funds are available for each indexed benchmark, even though the expected return of stocks and bonds is far more reliable an estimation than the expected return of a research program in astrophysics or paleoanthropology.

We propose here an alternative approach that considers the history of public R&D funding in the U.S. and seeks to uncover the logic behind it. We draw from the theory of punctuated equilibrium in public finance to submit the times series of the U.S. federal R&D, from 1976 to 2015, to scrutiny (the last year is an estimate based on the budget request to Congress). We first parse the analysis by type of spending—in the traditional basic, applied, and development typology—and then by federal department or agency—these are not synonyms but in the interest of parsimony, will use them here indistinctly. In this analysis we seek to identify how much of its long-term trend is accounted by incremental increases to the R&D budget and how much by budget leaps. We find gradual annual growth to be far more important than bursts and, in fact, we find the budget leaps do not have lasting effects. We then turn to a comparison of the historical behavior of agency R&D to the respective agency total discretionary budget, and we find that in most cases R&D tracks the total agency budget. This finding intimates a peculiar politics of the R&D budget that, we posit, leads agencies to adopt one of two dominant strategies. When R&D is a small part of the agency budget, as is the case of the DoD, agencies keep it as inconspicuous as possible. When R&D is the largest component of the budget, agencies instead self-promote insisting on the importance of their work to the national interest. In order to maintain long-term incremental growth, agencies either keep R&D invisible or elevate it to a high moral purpose. Both parts of the analysis suggest a relative inter-agency equilibrium and a growth rate that is dominated by incremental annual gains; not by leaps. We conclude returning to the motivating question and consider some policy options that take these findings into account for the federal government to set priorities.

2. THE HALF-LIFE OF PUNCTUATIONS

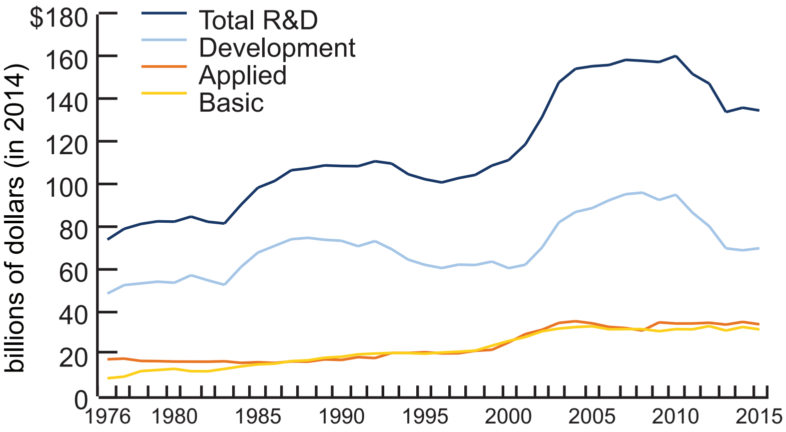
In order to analyze the historical evolution of federal R&D, we draw the statistical methods prescribed by the punctuated equilibrium theory of policy making. In policy studies, this theory explains two empirical characteristics of the evolution of policies, the incremental change that dominates most of their trajectory, and the shocks that policy systems experience from time to time.

Punctuated equilibrium in policy studies took its name and inspiration from biology's evolutionary hypothesis to explain "bursts" of evolution. The fossil record shows periods of stasis or very gradual adaptation in the evolution of organisms that are punctuated by evolutionary leaps. In policy studies, this theory has its origins in Charles Lindblom's famous essay *The Science of Muddling Through* (1958) where he set out to formalize the actual process of incremental decision making that takes place in the bureaucracy, in contrast to the comprehensive rational model that his contemporaries assumed. In public finance, Aaron Wildavsky (1984|1964) collected significant evidence in support of Lindblom's formalization. More recently, Baumgartner and Jones (1993) observed that the general tendency to incremental decisions is upset by intermittent punctuations spaced in time and characterized the political dynamics of this process. At the heart of their explanation is that equilibrium in a policy subsystem is sustained by the work of a policy monopoly (a coalition of interest groups) that benefits from the status quo; when this monopoly is undermined and broken, other political interests succeed in pushing for policy reform (see summary in True, Jones, and Baumgartner, 1999).

We borrowed from this theory the statistical test of punctuation in the federal R&D. While there is no single standard or ideal method (Jones, Sulkin, Larsen, 2003), and oft-used technique to discover the presence or absence of punctuations is to test the kurtosis of the underlying distribution (True, 2000; John and Margetts, 2003). The probability distribution of the variable in question is estimated from the histogram of first differences (we use the growth rate here). If most years the change in budget is around a mean change for the sample, the series does not often experience significant leaps. This scenario corresponds to a leptokurtic distribution; that is, with a very high concentration around the mean and thin tails—the thinner the tails the scarcer the punctuations. The standard kurtosis test has a critical value of three, which means that values statistically larger than three are associated with a leptokurtic distribution.

We can observe (Figure 1) that total R&D has a general upward trend characterized by incremental increases albeit with some volatility in its trajectory. Applying the kurtosis test to total federal R&D and its traditional components—basic

Figure 1. Federal R&D 1976-2015



Source: AAAS compiles R&D data based on US Office of Management and Budget publications; see http://www.aaas.org/page/historical-rd-data.

research, applied research, and development—we do find that gradual adjustments dominate these time series (kurtosis is greater than three), yet punctuations are large enough for us to reject any statistically significant difference between the observed kurtosis and that of a normal distribution of first differences (Table 1) with exception of basic research.

It is worth noting that total R&D has had an average annual growth rate from 1976 to 2015 of 5.1 percent and over the same period annual inflation has been around 2.4 percent, suggesting that the federal R&D purse has experienced consistent purchasing power gains. Recall that as little as two percent annual growth above inflation doubles an agency's purchasing power in 35 years.

We realize that the moderate volatility of total federal R&D in the aggregate may not be indicative of volatility in each of the comprising agencies; rather, it may be affected by one or two agencies while the rest displays low volatility. Hence, we deploy the analysis of punctuation on each agency R&D time series (Figure 3 and Table 2).

We find that most agencies display either leptokurtic or normal distributions in first differences of their R&D budgets. This means that they have experienced a great deal of stability over the studied period; a period long enough to include economic upheavals, military engagements, the regular swings of the political pendulum, as well as changing priorities at each agency. The overall stability observed suggests that agency heads maneuver to implement new priorities within the limits of their overall budget—once the budget is capped, it is a zero-sum game within the agency—and indeed, it is rare that an agency would propose an overall increase in their budget in order to promote a single R&D initiative or program—the most notable exception to that rule is the Apollo program.

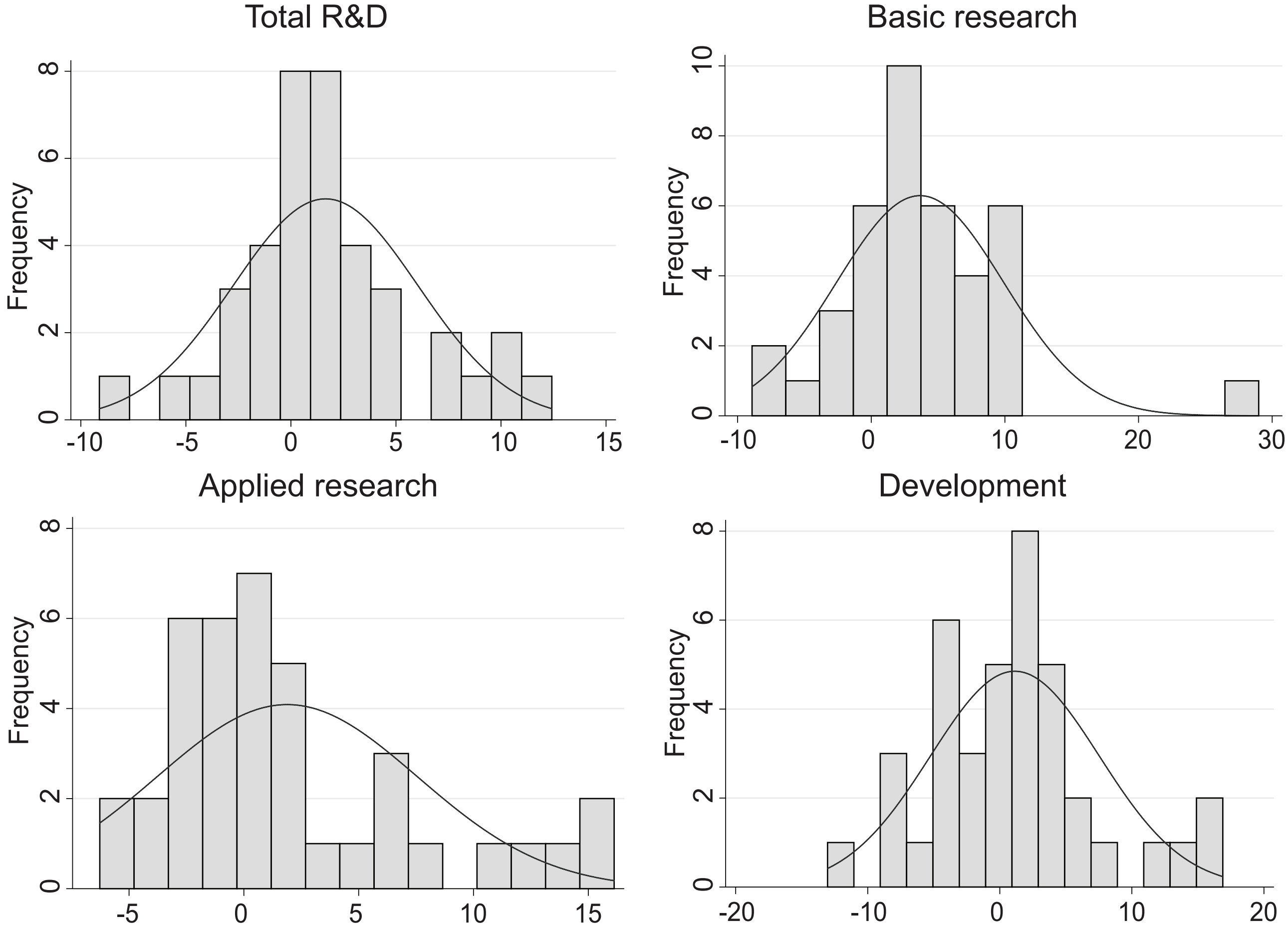
When this exercise is extended to first differences with three-year and five-year lags, the stability of the budgets increases even further and fewer leaps are observed than with only one-year lag. This finding further supports the earlier conclusion that overall federal R&D is stable over time; now we can extend this conclusion to the agency

Table 1. R&D components: Kurtosis of distribution of first differences

	Kurtosis		Skewness	
	Kurtosis Stat	Significantly Different From Normal	Direction of Skew	Significantly Different From Normal
Total	3.68		+	
Basic	8.28	**	-	
Applied	3.39		+	
Development	3.47		+	

Notes: P-value Level of Significance: *5%; **1%
Normal Kurtosis=3; Larger than 3 =leptokurtic; Smaller than 3=platykurtic

Figure 2. Federal R&D: Frequency distribution of first differences, 1976-2013



level as well. What is more, this finding may provide some solace to the research community insofar as R&D does not become an object of political wrangling.

We parsed the analysis by agency suspecting that one or two agencies with a relatively large and volatile R&D could be upsetting the statistical significance of our leptokurtic results at the total level. This seems indeed to be the case, and our analysis has revealed the culprits: defense and medical R&D. The kurtosis statistic for DoD and DHHS is indeed below the critical value of three (Table 2), and while it is not statistically different from the normal distribution, it does intimate that punctuations are not negligible for these agencies, and what's more, they can affect the total R&D level. This is not at all surprising given that they take the lion's share of federal R&D, with DoD consistently taking half of the federal total and the NIH (that represents above 95 percent of DHHS R&D) taking above a third of civilian R&D (Figure 4). It should be noted that the NSF is the third agency with a low kurtosis statistic but its punctuations hardly move the federal R&D needle because of its small size relative to NIH and DoD.

For most agencies then, their R&D trajectories are rarely punctuated. Requests for a substantial increase in an agency's R&D are rare and successful increases are even rarer. The celebrated work of John Kingdon (2010|1984) gives us a guide into the unusual set of circumstances that must come together to open a "window of opportunity" for policy to change: a widespread sense of urgency on the matter, a credible and well-debated policy proposal, and

Figure 3. R&D by federal agency: Frequency distribution of first differences 1976-2013

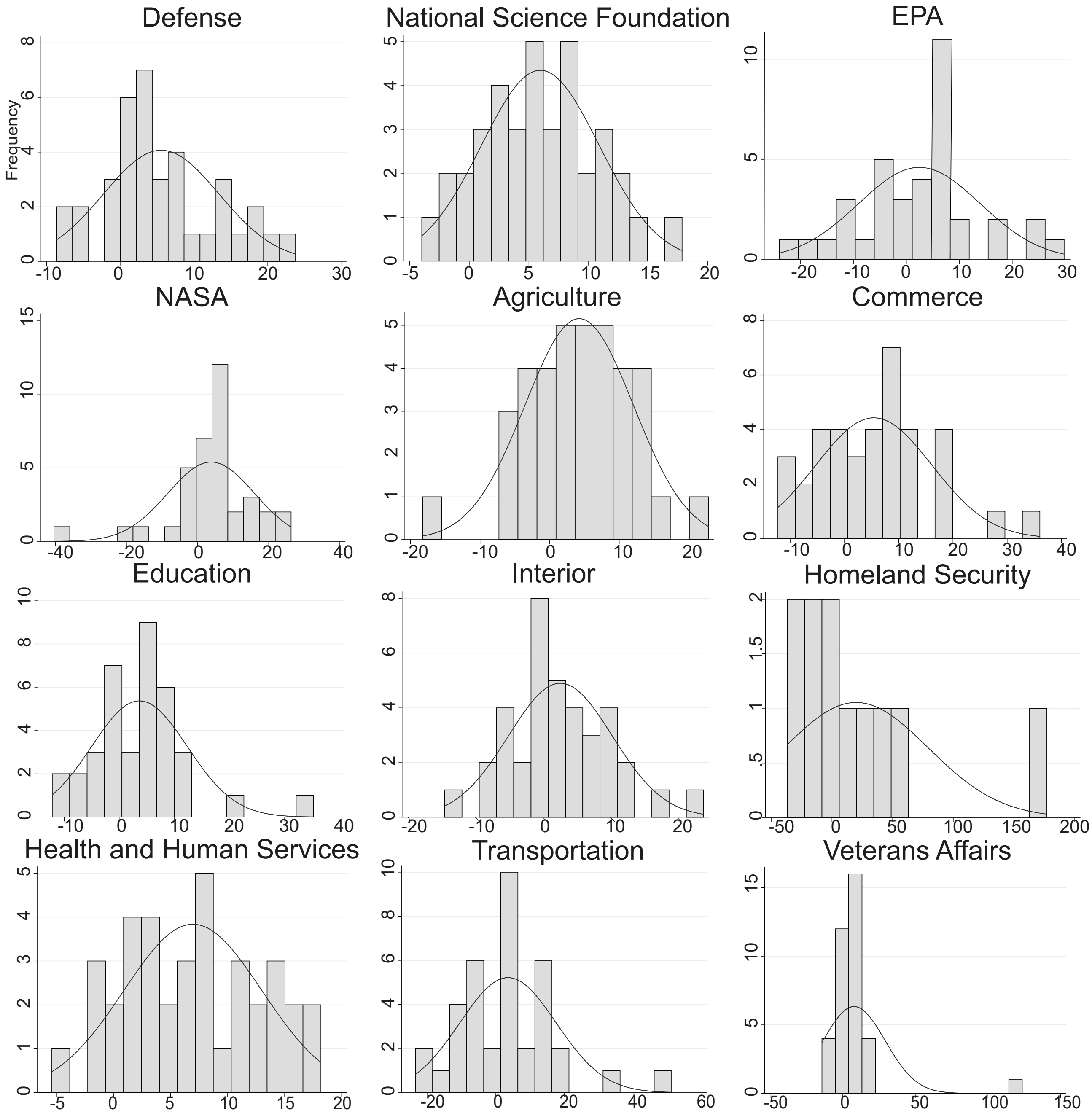


Table 2. R&D components: Kurtosis of distribution of first differences

	Kurtosis		Skewness	
	Kurtosis Stat	Significantly Different From Normal	Direction of Skew	Significantly Different From Normal
DOD	2.71		+	
NASA	6.54	**	-	**
DOE	6.30	**	+	**
HHS	2.15		+	
NSF	2.57		+	
USDA	3.73		-	
Interior	3.43		+	
DOT	5.13	*	+	*
EPA	3.25		+	
DOC	3.54		+	
DHS	5.58	**	+	**
VA	24.45	**	+	**
Other	17.59	**	+	**

Notes: P-value Level of Significance: *5%; **1%
Normal Kurtosis=3; Larger than 3 =leptokurtic;
Smaller than 3=platykurtic

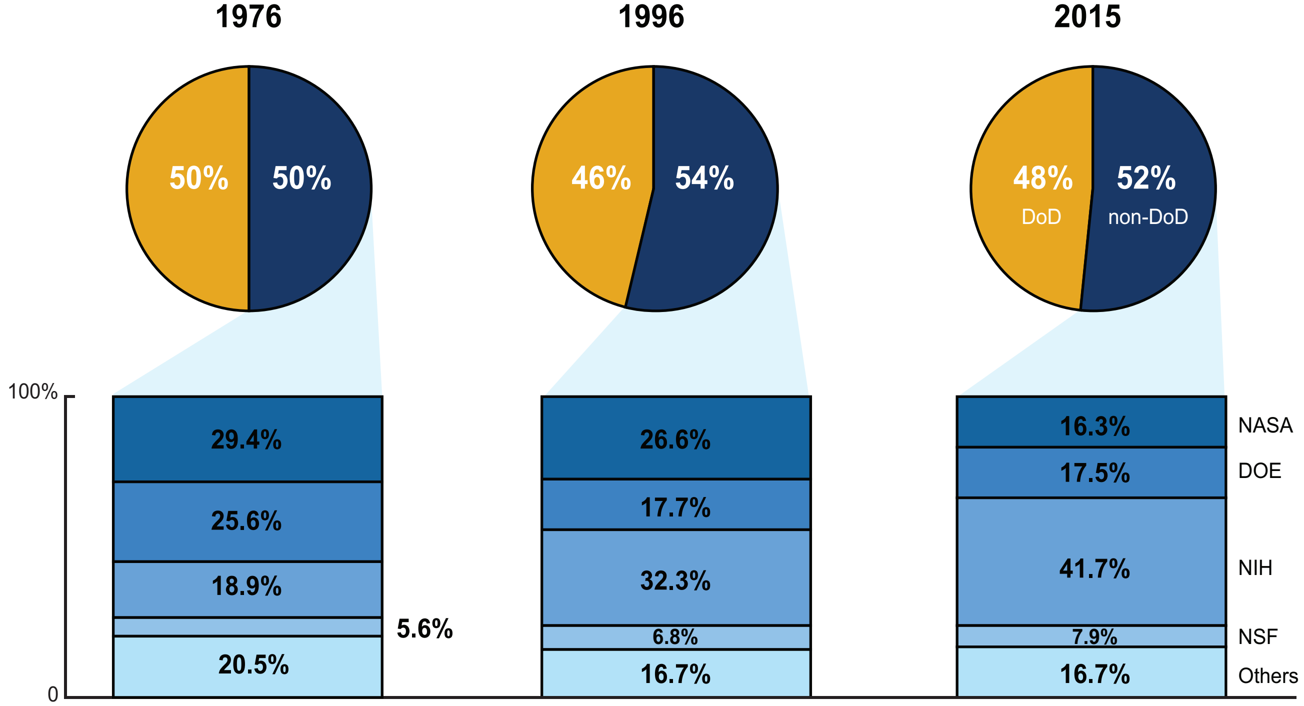
political will to produce the change. The opportunity will materialize into change by virtue of the diligent offices of a policy entrepreneur with the energy and determination to convince the White House of the political gains of the proposal, to outmaneuver the Office of Management and Budget at its game, and most improbably, lining up support while neutralizing opposition in Congress.

One surprising aspect of the stability in agency R&D budgets is that the few punctuations observed do not seem to have lasting effects; rather than producing a structural change, budgets seem to return eventually to their long-term trend. A notable example of this phenomenon is the doubling of the NIH budget from 1999 to 2003. Since this punctuation, the NIH purchasing power has stagnated (i.e. using a healthcare sector price index) and is now approximating the inflation-adjusted budget levels that preceded it. It seems that every time an agency manages to attract the favor of policymakers, it cannot retain it for too long. Agencies spend political capital in R&D faster than graduate students spend travel grants.

A distant example of punctuation that is nevertheless illustrative of the ephemeral character of political support for an R&D initiative is the Apollo program. We hear from time to time calls for another national project of such a magnitude and sense of purpose, including energy (Obama's "all of the above" energy strategy), health (Nixon's War on Cancer), or once again space exploration—with extravagant missions such as the colonization of the moon or manned exploration of Mars. In spite of the fact that NASA succeeded in landing the man of the moon, not once but six times, it was not able to justify the same level of spending indefinitely. The Apollo punctuation is not only a story of technological marvel but also one of bureaucratic prowess because the remarkable technical achievement was only possible by virtue of remarkable organization. Still, Apollo is also a story of the short life of political support that made possible such a scientific and technological achievement.

The basic conclusion of this analysis is that, most of the time, for most agencies, the R&D budget will only increase incrementally. Policy shocks like the Apollo program in the 1960s or the doubling of the NIH budget in 1999-2003 do not seem to place the agency budget on a new trajectory. If gradualism is dominant in the story of agency R&D budgets, we should then turn our attention from the drivers of punctuation to the drivers of equilibrium in the time series; what makes the R&D crawl only slightly faster than inflation? We need to consider R&D spending in the context of each agency's total budget.

Figure 4. R&D composition by department/agency



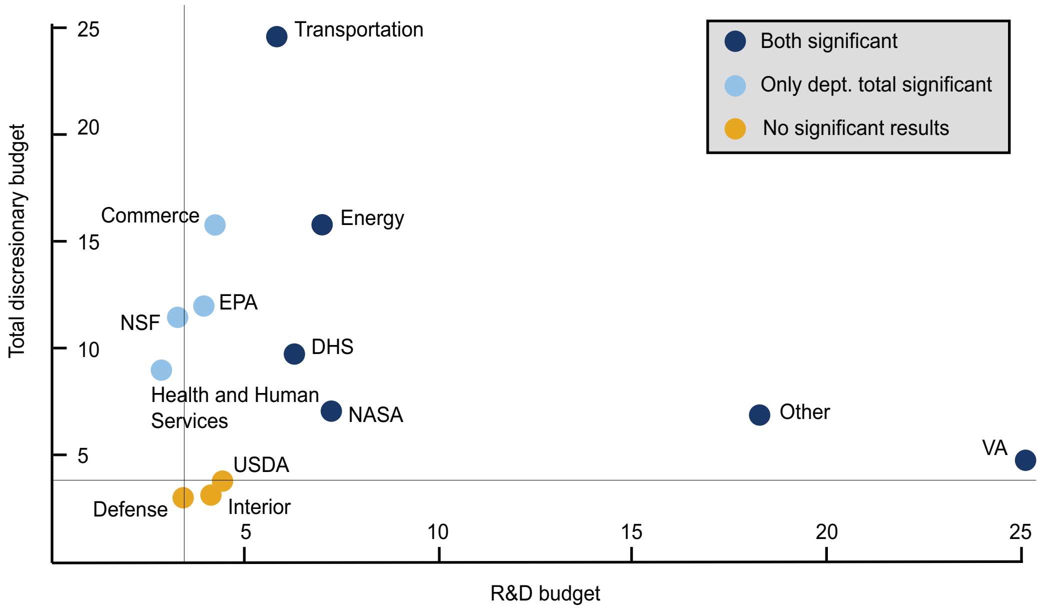
3. PROTECTING THE EQUILIBRIUM: CONCEALMENT OR SELF-PROMOTION

We want to examine whether the fate of R&D is tied to the fate of the total budget of the agency. The total budget that we will consider is only the discretionary part of total spending because decisions at the agency level are typically made on a year by year basis almost exclusively on the discretionary part of the budget. The nondiscretionary or mandatory spending is required spending by existing law (entitlements, interest, and debt payments) and changes to those programs are difficult and infrequent. Discretionary spending is where the action is in budget negotiations at the agency level.

We begin by testing the prevalence of punctuation on the total agency or budgets. Below we present the results of the test of kurtosis on total agency budget paired with the results obtained for agency R&D (see Figure 5). It becomes readily apparent that, in most cases, the rate of annual change has a leptokurtic or normal distribution in both time series, R&D spending and total spending for most agencies; which suggests that neither R&D nor total agency budgets experienced significant shocks in the period studied.

To be precise, Figure 5 displays the results in four quadrants arranged in colors that correspond to the four possible combinations of a two by two matrix, on each axis the two variables (R&D vs. total discretionary budget) are grouped in two categories determined by whether it meets the test of leptokurtosis with or without statistical significance (at 95 percent confidence). In general, the higher the value the likelier it is to be leptokurtic with statistical significance. The orange colored dots are associated with departments that have no statistically significant evidence of leptokurtosis for either their total budget or R&D budget. The light-blue dots are departments that demonstrate evidence of leptokurtic distributions for the total discretionary budget, but not for the R&D budget. The blue dots represent

Figure 5. Kurtosis test on R&D and total agency budget



Notes: North and West of the dividing lines show leptokurtic distribution of the first difference in the time series (above critical value for the test = 3).

departments that have statistically significant leptokurtic distributions in both total and R&D budgets. No departments in this study have leptokurtic distributions only in the R&D budget.

We had noted earlier that DoD, HHS, and NSF are the only agencies falling short of the critical value in the test of kurtosis (albeit neither with statistical significance); this intimates that at least some of their annual total budget changes were large enough to produce thicker tails in the respective probability distribution of first differences. We will return to discuss how these agencies have different budget politics, but it is worth noting that they are an exception—albeit, an important one—to the general rule that agency R&D budgets exhibit negligible year to year changes.

The absence of regular shocks in both R&D and total agency budgets suggests that R&D at each agency might be pegged to the total discretionary budget of that agency. Consequently, we examine the ratio R&D to total discretionary expenses for each agency—henceforth "R&D share"—to determine whether it has significant variability.

Agencies can readily be grouped into two distinct types: those where research is the main agency mission (high R&D share) and those where R&D is simply one of several functions and is not large relative to the size total agency budget (low R&D share) (see Figure 6 and Table 3). This will play a substantial role in differentiating the political strategy of the department or agency when seeking to secure an annual increase in their budget. We can also appreciate a few agencies have a more volatile R&D share than others—NSF, NASA, DoE (see Table 3)—due primarily to the fact that R&D is more stable than a widely variable total agency budget. This occurs only in agencies dedicated to research while low volatility is more common in agencies with a large R&D share.

A good example of low and stable R&D share is defense. DoD overshadows any other agency in R&D spending; as we noted earlier, the spending of all other agencies combined is comparable to that of defense alone (Figure 4). This is in large part the result of inertia in a policy trajectory that began in World War II. During the war mobilization the government first committed large spending levels in R&D and during demobilization it committed to sustain high

Figure 6. Average R&D share for two types of agencies

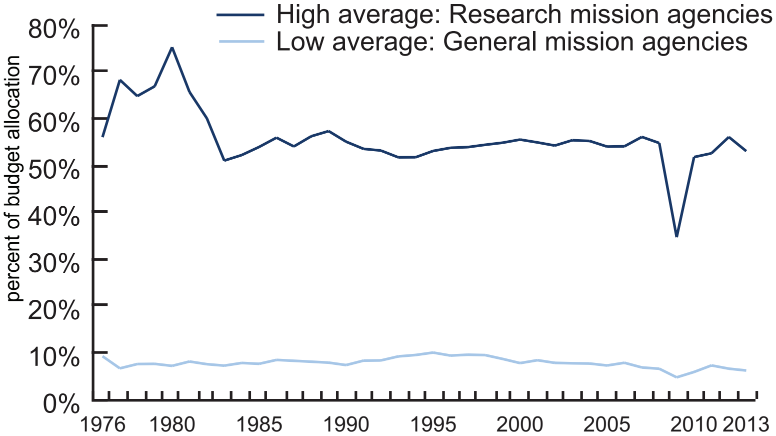


Table 3. Agencies with high and low R&D share

Dept.	Mean	Std. Err.	[95% Conf.	Interval]
High R&D Share Agencies				
NSF	81.2	1.7	77.8	84.7
NASA	65.6	2.6	60.3	70.9
DoE	41.9	1.5	38.9	45.0
NIH*	35.9	0.7	34.4	37.3
Low R&D Share Agencies				
DoC	17.8	0.8	16.0	19.5
DoD	12.5	0.3	11.9	13.0
USDA	10.4	0.2	10.1	10.8
EPA	8.1	0.6	6.9	9.3
DoI	7.4	0.2	7.0	7.8
DoT	5.0	0.2	4.6	5.5
VA	2.1	0.1	1.9	2.3
DHS	0.8	0.3	0.2	1.3
Other	0.7	0.0	0.6	0.8

Note: Each agency has 38 observations.
*NIH is a percent of DHHS discretionary spending. Non-NIH is only about 2 percent of DHHS'S R&D.

levels of R&D. In fact, the institutional basis for contemporary science and technology policy were established in the aftermath of the war.

During the Cold War, the U.S. confronted an existential threat from beyond the Iron Curtain and significant investments in defense were justified, including the pursuit of technological superiority over the Soviet Union. While the Cold War ended in 1991, the War on Terror that started in 2001 reactivated the national security imperative of high defense spending. In addition to commanding a muscular budget, punctuations in the DoD budget generally correspond to the beginning and end of large military deployments, notably Korea, Vietnam, Kuwait, and the Balkans.

It is interesting to observe that every time the Pentagon is funded to support a new military mission, R&D receives a boost that is not quite proportional but that is nonetheless substantial. In principle, there is a lag between R&D investments and their applications in field operations and no apparent reason why current increases in DoD budget need to increase current expenses of R&D. What happens—and this dynamic is central to our general explanation—is that R&D is not a separated function with a separate budget line. Rather, it is interspersed with several functions in the various armed forces. A new military mission expands the resource demand for deployment and supporting functions across all departments, and R&D rides along. All boats docked at the DoD marina rise with the mission tide.

Other agencies have a similar experience when these conditions are satisfied: that the agency budget is an aggregate of several subunits, that R&D is not the most important function of the agency and its share is relative small, and that several subunits perform R&D; in other words, R&D is dispersed, fragmented, and thus inconspicuous to budget negotiations of

B | Center for Technology Innovation at BROOKINGS

the agency. Advocates of R&D at the DoD, the Department of Transportation (DoT) and Veteran Affairs (VA) have thus an interest in keeping R&D invisible during budget negotiations.

Still, four agencies are primarily devoted to research, to wit: NSF, NASA, DoE and the NIH (within the Department of Health and Human Services, DHHS). Other departments are not primarily involved in research but have a substantial portion of their budgets devoted to it, such as the Department of Commerce (DoC) that manages the National Institute of Standards and Technology (NIST) and the National Oceanic and Atmospheric Administration (NOAA). Advocates of R&D in these departments and agencies, in stark contrast to their counterparts at DoD, DoT, and VA, argue for their budget primarily on the merits of R&D.

In summary, overall federal R&D statistics are stitched together given that each piece is managed separately. In the decentralized budget process, there is no systematic method guiding R&D allocation. Our portfolio manager is either idiosyncratic or worse, frenzied. If the test of federal policy is to manage R&D like a mutual fund, Daniel Sarewitz (2007) is correct in wondering whether science policy really does exist in the U.S. The absence of an allocative logic should not be confused however with the absence of a general strategy that agencies follow to protect their R&D budgets. Two such strategies are dominant. Agencies either conceal R&D in their item lines or highlight its importance to the national interest; it depends of the relative size of R&D in their budget and its prominence in the agency's mission.

We obtain two corollaries from this analysis. First, the relative social importance of an agency's R&D is not a precondition for their public salience. That the funding of pure science or biomedical research are regularly and loudly defended in the public forum is determined more by how important federal funds are for the NSF and NIH than how important federal support in these areas is for the nation. For clarity, we are not suggesting these particular research investments are unimportant to society, but that we cannot infer their relative importance to society by how visible they are. The policy corollary is that changing the relative importance of R&D within an agency could change its politics because it forces that agency to sell its budget request with a different pitch.

4. NAVIGATING THE NEW POLITICS OF R&D BUDGETS

At the outset we asked how the federal government ought to assign R&D priorities in the changing political environment. We learned that R&D budget decisions are decentralized and that there is no central allocative logic other than agencies maneuvering independently to follow a path of incremental changes. Our question is then indirectly answered in disaggregated form: Priorities are assigned by the individual efforts of agencies to protect their R&D lot. We characterized these efforts under two general strategies, concealment and self-promotion. In this context, our original question amounts to asking what agencies should do in response to the new budget politics.

Political polarization, in budget terms, could be deemed a victory for fiscal conservatives because it will likely keep the growth of fiscal deficits low. The period we have examined here, has been generally one of expansionary fiscal policy, but we have now entered into a period of stagnant budgets. Overall federal R&D will likely level and the best-case scenario for federal agencies will be that they get to keep their R&D purchasing power at current levels. The worst-case scenario is for agency R&D budgets to be targeted for cuts. While in the past the champions of R&D could try from time to time to propose substantial budget increases for their respective agencies, they will have to become more bearish and develop strategies to fend-off the threat of cuts to their R&D. In addition, tighter budgets invite greater scrutiny from congress and calls for greater accountability. Congress will reward the audacity of special requests for additional R&D funding with greater scrutiny.

In short, the new politics of the R&D budget will discourage agency requests for special budget treatment and will increase accountability pressure on agencies and research organizations. Research agencies and organizations will be well-advised to adopt two postures in response: modesty in their budget ambitions and caution in the semantics of accountability. First, they should emphasize the importance of budget stability in the long run, even for very gradual increases; this also implies that they must abandon any projects to promote leaps in their R&D budget. Second, agencies and organizations must carefully adopt the new language of accountability and the corresponding mechanisms of implementation.

Regarding our first recommendation, we believe that R&D intensive agencies should pursue sustained gradual growth over temporary jumps in R&D funding. This advice is motivated by avoiding disruption in the research enterprise. Research, more than other social functions, uses very long time horizons for planning; the stability and predictability of its budget in the long term is critical for its smooth operation.

We have shown thus far a stubbornness of the long term trend in federal R&D at the aggregate and agency levels. This is a strength rather than a problem because it is a side effect of the resilience of the U.S. innovation system. Conversely, R&D in the private sector is sensitive to the business cycle because it is funded with gross profits; hence, it is critical for the U.S. to maintain a base-level of scientific and technological effort that is unaffected by economic downturns.

This base-level is doubly important to sustain for long periods of time when considering the high costs of building human and physical capital specialized in research. The training of scientists and engineers in the conduct of research demands substantial time and resources—many of which are public—with the result that other occupations for this labor force are only a second best use of their skills. In addition, once a research career is interrupted for a period of time, the highly competitive environment makes it very hard to hop back on the track. Similarly, laboratory buildings and equipment are not easily retooled, if that is at all possible, for the production process. Investments in research capacity are often sunk costs.

Investments in the education of students pay off years later when the students complete their training. Any significant downturn of the R&D below a base-level would force entire cohorts of researchers to seek nonresearch jobs. Likewise, any significant upturn would result in overloading the pipeline today driving an excess supply of research at some point in the future, and when the punctuation recedes back to the trend, new graduates will find themselves fighting a cut-throat competition for a limited number of research positions—or extending their post-doctoral training indefinitely. This is the experience of many scientists in the biomedical workforce who began advanced training at the time of the aforementioned doubling of the NIH budget and entered the job market by the end of that punctuation; currently, taking several years of post-doctoral training is the norm among PhDs in the relevant fields.

These are strong reasons why the goal of champions of R&D should be to guarantee a gradual but sustained rate of growth for their agency R&D rather than pursuing bursts of funding. In addition, tighter fiscal budgets are driving policymakers to be more curious about the uses of public moneys in research. Any future budget request above and beyond gradual growth will invite additional oversight and scrutiny; already legislators are taking a closer look at the decision making process of federal agencies in allocating their R&D funds. Policymakers are keen about a particular notion of accountability that insists on performance measurement and quantified causal links from public inputs to organizational outputs to societal outcomes.

Our second recommendation to agencies funding and organizations performing R&D concerns these aspects of the new politics of the R&D budget, to wit: increasing accountability pressure and the expectations built-in the language of accountability.

In the introduction we suggested that policymakers are using the language of returns of investment to hold federal agencies accountable for their R&D spending. ROI is used as well in policy discussions on the economic impact of research conducted by universities and national laboratories. At the outset, we also stated the reasons for our skepticism that ROI is useful to inform federal R&D priorities and while we recognize its growing currency, we believe that this tendency must be resisted. We offer an additional reason here; a reason that may ally in its cause even advocates of ROI estimates, put briefly, by equating accountability to performance or outcome measurement is a strategy that guarantees the disappointment of legislative overseers and the frustration of administrators of agencies universities, and laboratories.

Edwin Mansfield produced the first estimate of the economy-wide ROI of research in 1977 (Mansfield, 1977; 1998). The figure of 28 percent is quoted far more often than the careful caveats he included about the limited use of this estimate for policy. One notable exception is a Congressional Budget Office (1993) memo that highlights that “the very nature of the estimating methodology, as Mansfield has noted in his articles, does not lend itself to use in the annual process of setting the level of federal investment in R&D, nor to allocating that investment among its many claimants” (CBO, 1993). Estimates of the return of R&D on GDP can overlook the internal complexity of the system because the calculation is done at the aggregate level and because interactions, spillovers, and other indirect and mutually cancelling effects are thus pooled together. When we start looking for ROI in subsystems, we begin to lose account of complex interactions and begin to confront the intractability of linking causally different units of analysis as is the case of research at the agency or organizational level and economic activity at the regional or national level. The challenges with estimating the ROI of research at the subsystems level is only one case of the more general problem of mapping causally inputs, outputs, and outcomes. Any model that links research—at the agency, organization, program, or grant level—to outcomes will make very far-reaching assumptions to simplify the complex way in which research mixes at several points in the production of the specified outcome, not to mention a lack of specificity of the outcomes of interest. Estimates from such models will be highly sensitive to assumptions and thus will be hotly debated.

The testing of hotly debated alternative hypotheses is what keeps science ticking. Hence, the search for quantification of research outcomes and the precise mapping of inputs to outputs to outcomes is a laudable intellectual enterprise that must continue. Serious attempts at quantifying inputs, outputs, and outcomes, such as those of the STAR METRICS consortium (Lane and Bertuzzi, 2011) are worth considering not only for what they achieve but for the insurmountable challenges they uncover. But this line of research must not oversell its promise to policymakers that it will settle policy debates about the productivity of research. The measurement of performance cannot expunge the politics from public accountability. Nor it can eliminate uncertainty. In fact, measurement—in the absence of a broad consensus about the legitimacy of the measuring rod—creates a political problem of its own: any estimate used by proponents of a research program will be denounced as inflated by detractors of that program. It creates an analytical problem as well, because an adequate representation of uncertainty (such as a range of equally likely estimates within a confidence interval) will undermine the ability of the estimate to council action (e.g. what to do with a program with a ROI of 3 percent that is just as likely to be -5 percent and +11 percent). Since estimates of long-term returns are so sensitive to assumptions—such as calibration of initial conditions, modelling of spillovers and externalities and subsystem interactions—any honest estimate can have a range of probability so wide as to support opposite decisions at the same time (e.g. to fund and to defund a program) or any two fair-minded and credible analysts can produce estimates sufficiently different to feed opposite political appetites.

Practical stewards of R&D, particularly those at the helm of research agencies, universities, and laboratories, will do better avoiding committing themselves to the language of returns and performance measurement. Instead, they should emphasize evaluation by examining the integrity of the research process and their efforts to make this process responsive to public values and social expectations about research. However, measurement is the common currency of our days and cannot be wished away. There are two parallel tactics that can effectively be implemented to manage this reality.

The discussion of outcomes need not pursue the unrealistic aspiration of extricating politics from the budget process; instead, it can pursue the framing of the political debate by embracing not abstracting complexity, which means among other things, to offer a richer description of selected examples that do link research at the subsystem level to societal outcomes. Contrary to stripping understanding from complexity—as models do—policy debate benefits from a nuanced understanding that is brought to bear by thick description. Of course, an agency cannot produce a short history for each and every grant, but it can do so for a small sample carefully curated to illustrate the range, the complexity, and the adequate time horizon to assess its research. A thick description of outcomes can take the form of an examination of how a program or grant advances public values, both by design and by accident. The mapping of public values is an approach already in practice that has yielded valuable descriptions of the outcomes of research (Bozeman and Sarewitz, 2005; 2011).

The second tactic is to encourage outcome measurement but only within its proper sphere of application. We said that mapping research to outcomes is marred by intractability because research outputs are only vaguely specified—publications or patents—they take a long-time to realize their promise and they do so only after interacting with a myriad of other factors and efforts which altogether produce an outcome. In turn, most development contracts have a much better specified output than research—in terms of the expected characteristics, operational standards, and uses of the output—therefore, development outputs are far easier to map onto outcomes. For example, if the outcomes of interest are the economic value added and the jobs created in the production and commercialization of a new product, the impact of original research that contributed one patent among hundreds of patents used in the product is much harder to estimate than the impact of the prototype of that product. The second tactic, stated without nuance, is to emphasize process assessment for research and confine outcome measurement to development in R&D.

5. CONCLUSION

We have found that the upward trend in the federal R&D budget over the last four decades is explained more by gradual increments than punctuations, both at the aggregate and the agency level. Each year, the R&D budget of departments and agencies is largely determined by past levels and that has placed it on a relatively stable trajectory over time. The budget process is decentralized and does not pit agencies against each other in a race for a bigger slice of the R&D pie; each agency requests an R&D level looking at its past, not contrasting its merits against other agencies. It is useful imagining a different arrangement in order to highlight the politics of the current system. If the process would instead be centralized in a governmental office responsible for setting the level of R&D and allocating funds each year among the various federal agencies, it is likely that we would observe far more volatility in the historical series of each agency's R&D time series because allocation would reflect changing priorities and political moods. We would also observe a different budget politics as agencies would compete with each other for R&D funds; they would have to argue not only for the merits of their own programs but also for their superior social impact compared to other programs.

Taking a budget perspective, we also argued that agencies are of two distinct types: those where R&D is a significant part of their budget and those where R&D is only one more function of the agency. Agencies with small R&D units, simply let R&D be inconspicuous within their total budget; in turn, agencies primarily devoted to R&D extol the virtues of their programs and their positive impacts on society.

However, the politics of the R&D budget are likely to change. Political polarization is likely to drive a very slow growth of the fiscal deficit (without raising taxes) and perhaps even cuts across the board if sequestration is not revoked. We have already seen, over the past few years, tighter budgets and a concomitant increase in accountability demands from legislators onto federal agencies and research organizations. This pressure has been operationalized in the language of returns on investment and more generally, the formal quantification of causal links from public R&D taken as input, to organizational outputs, and to societal outcomes. We have made a case that this may be a more apt approach to evaluate development in R&D, and that agencies should stress that research takes a very long time to realize its promise and with few exceptions cannot be linked through neat causal chains to outcomes.

It is important to underscore that the case against equating performance measurement to public accountability is not a case against demanding public accountability from the research enterprise. Quite the contrary, we believe that publicly funded research can and should improve its responsiveness to public values and social expectations about research, and we have suggested alternative methods of evaluation. Qualitative assessments allow for a more adequate consideration of the complex links between research and outcomes over appropriate time horizons. They also place greater emphasis on the integrity and responsiveness of the scientific process. Qualitative evaluation can also integrate quantitative indicators when appropriate and carefully interpreted—this is in fact a recommendation emanating from the scientometrics community in their recent Leiden Manifesto (Hicks, et al, 2015). But we should remain wary about conducting research evaluation primarily on performance measurement, lest we introduce perverse incentives and heighten moral hazard and thus undermine the very goal sought by greater accountability.

We have further suggested that in an environment of tighter budgets, asking for additional R&D dollars over a base rate will likely invite a boiling accountability pressure on the initiatives proposed; if taking to an extreme, we fear potentially valuable initiatives may experience "death by oversight". What's more, our data shows too that budget leaps do not seem to have lasting effects; R&D budgets tend to return in time to their old trajectory. Going forward, we therefore recommend agencies to focus on sustained but gradual growth of their R&D budgets.

We hope R&D does not fall further as a victim of partisan politics—and the recent House vote passing the America COMPETES Act is not encouraging. To protect public R&D from political moods and to preserve a base-level of R&D activity to support the various missions of the federal government, departments and agencies should seek to institutionalize gradual increases to their R&D budget. In the past, a general stability in R&D spending has been a fortuitous accident of the budget politics; in the future we may benefit from comparable stability but only by the offices of R&D champions in the federal government. These advocates will do well in abandoning calls to double their agencies' R&D budgets and embracing instead efforts to institutionalize sustainable incrementalism.

ACKNOWLEDGEMENTS

We thank Daniel Sarewitz for insightful comments as wells as the participants to the *Atlanta Conference on Science and Innovation Policy* and the 2013 meeting of the *Association for Budgeting and Financial Management* where we presented earlier versions of this paper. We thank also Joshua Bleiberg for excellent research and editorial support. Benjamin Clark would like to acknowledge support from the Graduate School of Cleveland State University.

REFERENCES

Baumgartner, Frank R., and Bryan D. Jones. 1993. *Agendas and Instability in American Politics*. 1 edition. Chicago: University of Chicago Press.

Bozeman, Barry, and Daniel Sarewitz. 2005. "Public Values and Public Failure in US Science Policy." *Science and Public Policy* 32 (2): 119–36. doi:10.3152/147154305781779588.

_____. 2011. "Public Value Mapping and Science Policy Evaluation." Minerva 49 (1): 1-23.

Congressional Budget Office. 1993. *A Review of Edwin Mansfield's Estimate of the Rate of Return from Academic Research and Its Relevance to the Federal Budget Process*. http://www.cbo.gov/sites/default/files/93doc174.pdf.

Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah de Rijcke and Ismale Rafols. April 23, 2015. Bibliomerics: The Leiden Manifest for research metrics. *Nature* 520: 429-431. Doi:10.1038/520429a

John, Peter, and Helen Margetts. 2003. "Policy Punctuations in the UK: Fluctuations and Equilibria in Central Government Expenditure since 1951." *Public Administration* 81 (3): 411–32. doi:10.1111/1467-9299.00354.

Jones, Bryan D., Tracy Sulkin, and Heather A. Larsen. 2003. "Policy Punctuations in American Political Institutions." *American Political Science Review* 97 (01): 151. doi:10.1017/S0003055403000583.

Kingdon, John W. 2010. *Agendas, Alternatives, and Public Policies, Update Edition, with an Epilogue on Health Care*. 2 edition. Boston: Pearson.

Lane, Julia, and Stefano Bertuzzi. 2011. "Measuring the Results of Science Investments." *Science* 331 (6018): 678–80. doi:10.1126/science.1201865.

Lawrence, Jill. 2015. "Profiles in Negotiation: The Murray-Ryan Budget Deal." *The Brookings Institution*. Accessed April 20.http://www.brookings.edu/research/papers/2015/02/profiles-negotiation-murray-ryan-lawrence.

Lindblom, Charles E. 1959. "The Science of 'Muddling Through.'" *Public Administration Review* 19 (2): 79–88. doi:10.2307/973677.

Mansfield, Edwin, John Rapoport, Anthony Romeo, Samuel Wagner, and George Beardsley. 1977. "Social and Private Rates of Return from Industrial Innovations*." *The Quarterly Journal of Economics* 91 (2): 221–40. doi:10.2307/1885415.

Mansfield, Edwin. 1998. Academic research and industrial innovation: An update of empirical findings. *Research Policy* 26: 773–776.

Sabatier, Paul A., ed. 2007. *Theories of the Policy Process, Second Edition*. 2nd edition. Boulder, Colo: Westview Press.

Sarewitz, Daniel. 2007. "Does Science Policy Matter? | Issues in Science and Technology." Accessed April 20. http://issues.org/23-4/sarewitz/.

True, James L. 2000. "Avalanches and Incrementalism Making Policy and Budgets in the United States." *The American Review of Public Administration* 30 (1): 3–18. doi:10.1177/02750740022064524.

Wildavsky, Aaron B. 1984. *The Politics of the Budgetary Process*. 4th edition. Boston: Little, Brown.

GOVERNANCE STUDIES

The Brookings Institution
1775 Massachusetts Ave., NW
Washington, DC 20036
Tel: 202.797.6090
Fax: 202.797.6144
brookings.edu/governance

EDITING, PRODUCTION & LAYOUT
Nick McClellan

EMAIL YOUR COMMENTS TO GSCOMMENTS@BROOKINGS.EDU