Policy Diffusion and Legal Impact in Executive Compensation Disclosure: Evidence From a Natural Experiment

David Glick
Post-Doctoral Fellow and Visiting Assistant Professor of Public Policy
Rockefeller Center for Public Policy
Dartmouth College
David.M.Glick@Dartmouth.edu*

September 20, 2010

Abstract

When, and how, private actors comply with laws and regulations are fundamental questions of governance and public policy. Flexible and complex laws and regulations force affected actors to make choices. These choices may diffuse like legislative policies. This paper uses recent executive compensation regulations as a natural experiment to investigate policy diffusion in implementation. This unique case avoids many identification challenges common in diffusion studies. The timing of compliance depended on the timing of previously scheduled proxy reports. This allows us to observe firms which had to complete their disclosures without the benefit of having others to observe, and firms that complied later, with opportunities to see what others had done. This paper uncovers a substantial effect of having models to follow when responding to flexible regulations. It also shows that the diffusion of compliance policies led to less compliance and reduced the law’s impact on behavior.

*David Lewis, Keith Whittington, participants at the Dartmouth American Politics Colloquium and the 2010 MPSA meetings all provided invaluable feedback. Thanks to Zoe Friedland and Gabi Tudin for their thorough and enthusiastic research assistance. I would also like to thank Jim Smith for sharing his CD&A expertise with me. All stylistic and substantive errors are mine.
1 Introduction

In 2007, the Securities and Exchange Commission (SEC) instituted new executive compensation disclosure requirements. The “Compensation Discussion and Analysis” (CD&A) rules were intended to increase transparency by requiring firms to publicly explain and justify why they paid their executives what they paid them in the previous fiscal year. In 2008, months before its sub-prime implosion, Lehman Brothers filed its first Proxy Report under the new compensation disclosure rules. Among other things, it reported that CEO Richard Fuld had been granted $146 million worth of long term restricted stock units (RSUs). Weeks later, former Lehman Brothers associate general counsel Oliver Budde sent a two page whistle blowing email to the SEC with the subject line: “Possible Material Noncompliance with New Executive Compensation Disclosure Rules.” Budde claimed that Lehman only disclosed two out of fifteen RSU grants and that the actual total was over $400 million dollars (Sterngold, 2010).

The Lehman Brothers example suggests three realities about these regulations. The first is that firms do fear public scrutiny of their pay practices. The fact that Lehman appears to have been constrained from disclosing all of Fuld’s pay implies that sunshine has the potential to affect behavior. The second is that compliance is not automatic. Assuming the allegations are true, this case is far from the first or last instance in which firms do not always comply with the law’s letter and/or spirit. The third is that a flexible requirement’s impact depends on policy choices and interpretations that firms cannot avoid making. In this case, firms had to make choices about what to include or how to include it.

When, how, and why private actors comply with laws and regulations are fundamental questions of governance and public policy. The CD&A regulations which Oliver Budde claims Lehman Brothers violated offer a unique opportunity to address questions such as “what determines compliance with the law?” Certainly factors such as the severity of sanctions and the threat of enforcement affect compliance. This paper focuses on another key mechanism. It uses the CD&A regulations as a natural experiment to cleanly investigate

---

1 He also claimed, consistent with academic estimates (Bebchuk et al., 2010), that Fuld dramatically understated his 2000-2008 earnings in his 2008 congressional testimony.
and measure the extent that policy diffusion affects the shape and magnitude of compliance. The policy diffusion literature has grown rapidly and major journals have published dozens of studies investigating and demonstrating the spread of policies from country to country, state to state, and city to city. This literature systematically understates the role of diffusion in public policy processes by focusing almost exclusively on diffusion and learning from others in policy creation. Diffusion is important in policy creation but it is also highly applicable to policy and legal implementation. In this case, private actors such as business firms are affected by laws and regulations and may have to make their own policy decisions about how to behave in response. They have to do so because sometimes by design and sometimes inadvertently, many important regulations and court decisions offer more flexibility than guidance. Given this flexibility to define compliance, understanding how affected actors decide how to comply is essential to studying the law’s impact.

This paper fills two gaps in the large and rapidly expanding diffusion literature while addressing broader “how” and “when” questions about legal compliance. One is the substantive gap that results from overlooking diffusion in policy implementation and leads to the understatement of diffusion processes in general. The second gap is methodological. While the idea that policy makers are influenced by others’ choices is quite intuitive and believable, data and inference problems have limited these studies’ ability to cleanly identify diffusion and diffusion mechanisms. This paper addresses both of these gaps by providing clean, quasi-experimental, evidence of diffusion in public companies’ implementation of executive compensation regulations. This case is methodologically advantageous because the timing of compliance depended on the timing of previously scheduled proxy reports. This attribute allows us to observe firms which had to complete their disclosures early, without the benefit of having others’ policies to observe, and firms that complied later, with opportunities to see what others had done. This design captures diffusion at work in implementation. The paper shows and measures the effect of having models to follow when responding to flexible regulations. Moreover, and perhaps most importantly, it shows that the diffusion of compliance policies led to less, and less substantive, compliance and reduced regulatory impact on behavior.

Investigating this natural experiment contributes three methodological advances over
other diffusion studies. First, I am able to observe the counterfactual “how similar would actions have been if they were independent.” I have been unable to identify another study in any context which does this. Doing so is essential to cleanly identifying diffusion and parsing independent adoption of similar policies from interdependent decision making (Volden et al., 2008). Second, I look at continuous and/or multifaceted choices rather than dichotomous “adopt or do not adopt” ones. Third, while other diffusion studies select policies which have already become common, this design allows me to observe policies which became common and those which could have, but did not. In this case, the “policies” that diffused resulted from choices that firms had to make to comply with the government’s rule change. While there are certainly differences between diffusion in this setting and diffusion in more conventional applications such as policy choice in legislatures, the executive compensation disclosure case supports a research design which enables much cleaner identification of diffusion effects than previous observational studies have in any context.

This study shows that firms’ responses to the law are indeed interdependent and that interdependence in compliance has real consequences for policy impact. It also estimates the magnitude of these diffusion effects while avoiding many of the confounds that limit other diffusion studies. It thus contributes to both the legal impact literature and more broadly to the policy diffusion literature across the social sciences. The fact that organizations learn from each other is not terribly surprising. It is consistent with many of our own experiences making decisions and with previous literature. What is novel is the clear way this paper identifies and measures these decision making tactics and their connection to the law’s bottom line impact.

The paper begins by briefly reviewing the literature on policy interdependence and diffusion. It then provides additional background about the CD&A regulations and their virtue as a natural experiment. Next, it describes the original data I collected. Before delineating evidence of diffusion, it quickly examines the experimental setup, pre-treatment balance, and random “assignment.” The diffusion analysis proceeds in three parts. In the first, I investigate two continuous variables, word count and readability scores which the SEC itself focused on in its initial review and comments. Next, I investigate attributes of these reports about which the SEC said nothing. These attributes are the number of voluntary tables
and figures, the way the report begins, and whether a firm provides a list of compensation peers. Third, I investigate whether these three choices were more correlated in the second group than the first. I find that policy choices within each report were more correlated in the second group which provides especially strong evidence of interdependence. Finally, after showing convergence in metrics like report length and the number of tables, I draw a more substantive link by showing that later adopters who had the benefit of observing others disclosed less, and less precise, information about the annual bonuses that they paid their executives. I conclude by suggesting ways that this natural experiment can provide leverage on other important questions regarding diffusion mechanisms, implementation, and executive compensation.

2 Diffusion in Policy Making and Implementation

Policy diffusion and related ideas have generated substantial and increasing scholarly attention across the social sciences. (For recent reviews see Dobbin et al. (2007); Elkins and Simmons (2005); Karch (2007)). Diffusion mechanisms, those that lead to choice interdependence and the potentially unnatural widespread adoption of certain policies, are both important and widely applicable. Mounting evidence suggests that diffusion processes belong with factors such as ideological preferences, electoral goals, and objective conditions in explanations of political actors’ policy choices. In legal and policy implementation cases, diffusion appears to rank alongside incentives, self interest, ideological preferences and other factors in explaining the actions that actors take to comply with, or evade, the government’s demands on them. Empirical policy diffusion studies in major journals have investigated a variety of organizational and substantive contexts. These include U.S. cities and states (e.g. Berry and Berry, 1990; Grossback et al., 2004; Shipan and Volden, 2006, 2008; Volden, 2006), nation states (e.g. Meseguer, 2006; Weyland, 2005), and business firms (e.g. Strang and Macy, 2001; Strang and Still, 2004). While much of this literature is empirical recent work has increasingly focused on delineating and parsing mechanisms. These mechanisms include information and learning (Bikhchandani et al., 1998; Braun and Gilardi, 2006; Meseguer, 2004; Volden et al., 2008), competition (Volden, 2002), and reputation and
legitimacy (DiMaggio and Powell, 1983). For example, Volden and Shipan (2008) attempt to measure and compare mechanisms such as learning, competition, and coercion in the adoption of smoking bans and find evidence of all of them at work.

This paper contributes to our knowledge of issues related to legal implementation, compliance, and impact and to the deepening and broadening diffusion literature. It fills gaps in both areas simultaneously. While much of the diffusion literature in political science focuses on policy creation in legislatures and elsewhere, the processes are also highly applicable to the non-governmental institutions on the other side of the ledger. In policy implementation, it is the firms, universities, hospitals, and other organizations that are making “policy” choices in response to the regulations and court decisions that affect them.

While some studies have highlighted legal flexibility and ambiguity while pointing to the role that diffusion and related mechanisms play in implementation, our understanding of diffusion and legal compliance is still limited. A few others have investigated and suggested connections between diffusion and legal impact questions based on factors such as the ambiguity and complexity of law, and the fact that it is often safer to have common practices (e.g. Barnes and Burke, 2006; Dobbin and Kelly, 2007; Edelman, 1992; Gould, 2005). These studies have shown diffusion effects in areas such as disability access, speech codes, sexual harassment policies, and affirmative action compliance using interview, survey, and observational data. The fact that diffusion mechanisms are at work in these areas implies that courts and regulators ultimate impact (e.g. Rosenberg, 1991) will depend on responses that propagate via interdependent responses to law and policy changes. If “compliant” responses spread, the law can have a substantial impact, but the opposite can happen as well.

Building on this previous research, this paper investigates, in detail, the role that diffusion plays in policy implementation. It tests the hypothesis that policy diffusion is an important mechanism in firms’ compliance with vague regulations. It investigates whether the concrete steps that firms take in response to the law will depend on what they learn from others in their industry. Additionally, it tests whether the opportunity to observe others’ responses affects how much firms comply with the letter and/or spirit of the law.

The other studies cited above have all provided some evidence of diffusion in implementation, but this paper does so much more cleanly. It does so by using a unique natural
experiment. Studying social learning with observational data is inherently challenging and rife with empirical confounds. Thus, this large and growing diffusion literature is subject to methodological limitations which may produce inaccurate estimates of diffusion effects and findings about diffusion mechanisms. These methodological limitations include conceptual issues, problems of identification and over-determination, and selection concerns. They apply to diffusion in both policy creation and implementation. This paper’s unique natural experiment and dataset helps overcome these challenges.

One crucial conceptual and methodological complication in this literature is distinguishing “diffusion” as a cause of commonality and “diffusion” as an outcome in which policies happen to become similar. As Elkins and Simmons (2005) explain, conceptually, the term “diffusion” has been used to describe both effects and causes. In the “effects” conception, diffusion is an outcome in which policies and practices have spread from unit to unit and become increasingly popular for any reason. In the “cause conception,” the term refers to mechanisms and interdependent processes that imply interdependent outcomes and correlation across units. In this conception, “the adoption of a practice by one actor alters the probability of its adoption by another” (Elkins and Simmons, 2005, p.36). Two neighboring states responding to similar external pressures could produce diffusion as an “effect” without diffusion as a “cause” for example. One more precise way to define this diffusion as a cause is that it is policy similarity that would not have occurred if actors could not observe each other. Two neighboring states responding to similar external pressures could produce diffusion as an “effect” without diffusion as a “cause” for example.

This lack of conceptual consistency is very closely related to the primary methodological challenge inherent in empirical studies of diffusion in all contexts. This empirical challenge is distinguishing policy diffusion and similarity that result from interdependent mechanisms from similarity that would have existed even if decisions were made without observing others’ actions (Volden et al., 2008). This empirical challenge is endemic to measuring social learning effects. It, along with other observational confounds, have prompted some to turn to the controlled laboratory environment to study diffusion (Tyran and Sausgruber, 2005).

This paper can address this methodological challenge, along with additional ones, in a way that previous observational studies cannot. The challenge is identifying policy conver-
gence that would not have happened if policy makers did not know others’ policy choices. The statistical method which dominates the field, Event History Analysis (EHA) (Berry and Berry, 1990), is not immune from criticism (Mooney, 2001) and its abilities to resolve the fundamental challenge are limited. These models estimate the probability that a policy maker adopts a particular policy, usually a yes or no dichotomous choice, in a particular unit of time as a function of internal and external factors. The external factors are supposed to capture diffusion as a process while the internal ones capture, or control for, the probability of independent policy adoption. For example, Berry and Berry (1990) introduced the method by studying state lottery adoption. This model included internal independent variables such as the state’s fiscal health and its level of religious fundamentalism. It also included an external independent variable, the number of neighboring states that had previously adopted lotteries.

These models are very sensitive to omitted variable bias and are only as good as one’s (and the data’s) ability to model the propensity to adopt policy X if acting independently (Fransese and Hays, 2007). Since neighboring states are often similar to each other and may have similar goals, the observational equivalence and identification problems may be particularly acute in exactly the places that others have looked for evidence of diffusion. If the analyst cannot account for all of the unobserved dimensions, shocks, and signals which may produce independent adoption of similar polices at similar times, the model cannot parse independent and interdependent causes. As Fransese and Hays (2007, p7) write, “inadequacies or omissions in specifying the non-interdependence components of the model tend, intuitively, to induce overestimates of the importance of interdependence and vice versa.”

This paper utilizes a natural experiment which avoids this problem. It’s design also avoids two other limitations present in nearly all, if not all, other diffusion studies. The first problem is a selection one. Other diffusion studies investigate policies that were known to have been widely adopted. Lotteries and smoking bans (Berry and Berry, 1990; Shiman and Volden, 2008) became fodder for “diffusion as cause” studies after the policies were known to have diffused in the “outcome” sense. This selection bias is not terribly problematic if one wants to understand how and why policy X became common and as many diffusion studies
do. On the other hand, many diffusion studies make more general claims about diffusion’s importance. Arguing that diffusion and/or any of its mechanisms are important factors in policy-making generally on the basis of studies only focused on policies that became common is problematic. Surely there are many policies that could have diffused, and shared traits with those that diffused, but did not. The current studies can tell us how likely it is that policy that a smaller city enacted was previously enacted by a larger neighboring one, but cannot tell us how likely it is that a given policy adopted in the larger city will spread to the smaller ones. These concerns are also highly relevant when trying to determine the importance of particular mechanisms and traits that increase the likelihood of diffusion. While it does not resolve these issues in a government policy-making context, this study does enable us to identify the set of plausible options in one case to see which became common and which did not.

The design can also help us reliably infer why those that became common became common. While this paper focuses the basic existence and magnitude of diffusion mechanisms, the natural experiment’s advantageous traits may help us infer and adjudicate diffusion mechanisms in the future. Finally, unlike most studies that focus on yes or no choices about a particular policy, this one investigates choices with continuous, or at least many, options. While this paper does not study particular mechanisms, the case’s advantageous properties and design offer substantial potential to advance our understanding of diffusion processes and mechanisms in future work.

3 CD&A Background

The CD&A regulations were announced in July 2006 and took effect in December. They affected companies’ annual proxy statements which preceded annual meetings beginning in January 2007. They were the first changes to executive compensation disclosure in 14 years. They were motivated, at least in part, by high profile compensation scandals including options back dating, unjustifiable severance packages, and lavish perquisites. As John White, the director of the SEC’s Corporation Finance Division said, “executive compensation has changed substantially since 1992 but disclosure rules have not, so chasms in disclosure have
resulted, and at the same time what the SEC did require has drifted toward boilerplate and legalese” (Institutional Investors, 2006).

The new requirements sought to increase transparency and accountability in executive compensation. Firms would have to explain and justify their compensation methodology, inputs, and decisions to shareholders, institutional investors, and the media. In theory at least, this sunshine would act as a deterrent to unreasonable and unjustifiable compensation. Proxy reports were supposed to include more comprehensive and straightforward tables, particularly the standardized “summary compensation table.” They were also supposed to include a “principles based” narrative in plain English to explain the compensation program and numbers. This second part was critical and is the focus of this investigation. The SEC wanted companies to describe their compensation programs in a way specific to their own circumstances and processes and not use standardized language or legal and accounting jargons. In an ideal response, a firm would figure out what language, figures, graphics and other information would best help shareholders and the public understand how its senior executive compensation process worked and how the firm arrived at the numbers in the Summary Compensation Table. These reports were deemed to have been “filed” rather than “furnished” which increased the board and executives’ liability for the disclosures when faced with enforcement actions, e.g. lawsuits over non-compliance.

The “principles based” approach that the SEC adopted is not unique to this case. Choices between “principles based” and “rules based” regulation will likely increase in salience as, for example, they will be highly germane to regulating the financial system. Moreover, many court decisions resemble “principles based” regulation as courts often lay out broad guidelines without specifying compliance details, and in both cases the potential for additional rounds of “feedback” and/or sanctions after the initial rule change persists. The principles based approach was meant to avoid uninformative boilerplate disclosure and language. In fact, SEC Chairman Christopher Cox told firms to start from scratch and create customized independent reports. In doing so, he more or less announced that this would be an interesting test case for policy diffusion. “The Compensation Discussion and Analysis is a brand new creation. Obviously no one has drafted these before. There’s no boilerplate out there. No precedents to mark and reuse.....Each company has a chance to start with a clean slate.”
(Public Comments 3/23/07). As we will see, despite the SEC’s calls for independence, the decisions that companies made to implement these regulations were anything but.

The regulations required firms to address six issues including, for example, compensation objectives and how the company determines the amount of each compensation element. The SEC also suggested a number of other things firms address including, the role of executive officers in the compensation process and whether compensation is benchmarked to peer companies and who these peers are. Figuring out how to include these elements, meet the requirements, and balance the conflicting demands for comprehensiveness and concision proved difficult. The SEC’s reviews of 2007 proxies resulted in letters to hundreds of companies citing problems including a lack of specificity and a lack of plain English. In a March 2008 *Wall Street Journal* article, Phred Dvorak cited both the companies’ confusion about the requirements and investors’ confusion about companies’ disclosures (Dvorak, 2008).

### 3.1 Advantageous Empirical Properties

This case’s attributes make it an ideal one for investigating how firms learn from each other, how they respond to complicated regulations, how policies diffuse, and how “principles based” regulations work in practice. The two most important properties both stem from the fact that the timing of compliance dates was a function of fiscal year and annual meeting dates. Since these dates are distributed throughout the year, some firms had to comply “in the dark” without any models to follow. Since these compliance dates were essentially fixed before the intervention, firms could not select into the early or later groups.

This case is unique because it allows us to observe a counterfactual that we usually cannot observe. We actually get to see firms “randomly assigned” to take actions with no models to follow. We get to see how similar or different their policies are from each others’ when they have to act independently. We also get to observe the full range of plausible options and actions that could have become popular rather than merely studying the ones that did. We then get to observe how later implementers’ policies vary when they do have an opportunity to observe what those in the first wave did. If anything, the findings will be conservative because even the baseline group was not fully independent. For example, legal and consulting firms were producing compliance guidance and recommendations which
might have induced more uniformity in the first group than true independence would have.

The second major benefit of using this natural experiment is that firms could not select into, or out of, the early adopting group. While in many cases this sorting is important and explains how lower capacity actors can do reasonably well when facing difficult decisions, it also confounds empirical analysis. For example, we may be interested in what traits make a policy diffuse and become common practice. One hypothesis is that firms are likely to copy prestigious industry leaders. If the industry leaders all act early and others do not, then leaders’ policies will be the only ones available to copy. The randomization in this case produces a variety of policy models from a variety of firms in the early period. It also produces a variety of firms remaining as later adopters. While I do not fully exploit this variation in this study, the design offers unique leverage on questions of who follows, who gets followed, and what makes a particular action diffuse. Future work can, and will, explore and exploit this uncommon and potentially highly valuable randomization.

4 Data and Design

One major benefit of investigating public companies’ SEC reporting is that much of the information is easily and publicly available. I first downloaded information from the Compustat Annual North American Companies Database for 2006 and 2007 via the WRDS site (Wharton Research Data Service). I collected basic company information (ticker symbol, industry etc.), financial information, and information about annual meeting and proxy dates for all public companies in the U.S. I then created the two sample groups from these data. To create the early proxy group, those who had to act independently, I took approximately the first two weeks worth of proxy filings from companies whose fiscal years ended on December, 31 2006. There were 160 companies with proxy dates in a two week period from February 27 to March 13th. These, along with the 13 other December 31st fiscal year companies that had earlier proxies (all but one in February), composed the early acting group. These proxy dates were so close together that these firms could not have learned from each others’ reports. The second group comprises companies with December 31st fiscal years but later (e.g. May instead of March) annual meetings, and firms with fiscal years that end
around the end of January. The proxies in the second group were released in May or June. There were 130 companies in the screen for the second group. I chose the second group so as to allow a couple of months for them to observe the behavior of the early adopters while minimizing the time for other conditions to change or other confounds to intervene. Even then, as I discuss later, I could not fully excise outside interference from the sample. The SEC began making public comments about the first reports in March.

Having created the two groups, I manually collected the dependent variables, elements of CD&A sections, which I found in the proxies in the SEC’s EDGAR web library of public filings. I decided which policies to collect based on reading a few of these reports and by soliciting suggestions from a friend who works for an executive compensation consulting firm. I collected the word count and readability data by copying all of the words in the CD&A, stopping at the “summary compensation table,” and pasting the contents (including tables and figures) into an online text analyzer. This tool returned a word count for each entry along with some readability scores (including a composite average). As I describe in detail later, I also collected data on the number of figures and tables in the CD&A, the heading on the first section after “Compensation and Analysis....,” and whether the firm disclosed a list of peers that it indexes compensation to. I collected exactly five dimensions of reporting and compliance and analyze all five below. The results do not reflect policies which happened to converge. They reflect the five policies which I decided to code without a-priori information about convergence. I also coded these reports’ disclosures of annual bonuses to assess the impact on an important area of substantive compliance. These data

\[^2\]I asked myself and my friend “what are observable and codeable choices that drafters had to make decisions about.”

\[^3\]This table is one of the new tables required and demarcated the end of the narrative part of the CD&A for nearly all of the proxies. In a few cases, the firm’s proxy looked different, and the summary table was either far from the CD&A narrative, or was mixed in with it at the beginning. In these few cases, I used my best judgment as to the boundaries.

\[^4\]http://www.addedbytes.com/code/readability-score/

\[^5\]In the process of doing this I had to discard some of my observations. In some cases the CD&A date in Compustat was inaccurate and the company did not belong in one of my groups. In others, a 2007 proxy was unavailable in Edgar for one reason or another.
are described in depth below.

4.1 Random Assignment?

“Assignment” to the groups was purely a function of proxy report dates which are closely related to fiscal years. Proxy dates change little from year to year. It certainly appears that assignment to the groups was random with respect to factors related to compensation and disclosure. The compensation programs and fiscal years would have been in place long before firms knew about the disclosures and certainly before they considered the possibility of being in the group that had to complete the disclosures independently. To verify, we can check that firms’ proxy dates were stable from 2006 to 2007. For the 225 firms that made it into Group-One or Group-Two in 2007, and for which both 2006 and 2007 proxy dates were available, the mean absolute change in days from one year to the next was 13, the median was six, and the mode was one. Seventy five percent of 2007 proxies fell within two weeks of the corresponding date in 2007. In other words, the composition of the two groups in the year before the regulations would have been very similar to the actual composition of the two groups in the first year of compliance.

Just because the groups were assigned randomly relative to the disclosures does not mean they were perfectly balanced. Fiscal years are not randomly assigned. For example, retailers often end their fiscal years in March because December 31st comes right at the end of their busy holiday seasons. The most relevant differences for these disclosures would be differences in the distributions of compensation consultants and industries between the two groups as both are likely connected to the mechanics of a firm’s disclosure. I will investigate these compositional differences in detail shortly when considering alternative explanations and confounds. In addition, the two groups vary significantly in dimensions like company age and market-cap. These differences, while real and which might affect executive compensation, are at least two steps removed from affecting the style and structure of executive compensation disclosure.

The best test of pre-treatment comparability is found in the companies’ own financial reports. I randomly selected 75 reports in each of the two groups. I compared the “audit committee report” sections. The audit committee report is another required section of the
annual Proxy Report but one which should be unaffected by the Compensation Discussion and Analysis rules. I ran each audit committee report though the text analyzer which counted the words and calculated a “readability” score as described above. *The two groups’ audit committee reports were indistinguishable.* They were the same length (the mean length was 455 words in Group-One and 434 words in Group-Two: p = .47), and had the same average readability scores (18.2 in Group-One, 17.9 in Group-Two: p=.42). These non-differences verify the quasi-random design. The reports are virtually identical across the two groups on a related but completely independent dimension.

### 4.2 Measuring Diffusion: Analytical Plan

The first thing this design will allow us to observe is the baseline level of policy variation when actors are given a vague directive and have to act relatively independently. We see this in the distribution of policies in the first group. Secondly, we can compare this first distribution to that of actors who complied a couple of months later. The analysis is quite simple. There are a few ways we can operationalize diffusion and we should see these in the data. If policies diffused, and those in the second group learned from the CD&As that they observed in the first period, we should see convergence through a reduction in policy variation. We are looking for convergence toward more common options in the second group relative to a wider variety in the first. The increased similarity in the second group can be interpreted as the causal effect of having opportunities to observe others. While we may see shifts in the central tendency as well, it is decreases in Group-Two’s policy variation relative to the baseline variation established Group-One that will indicate interdependence.

There is an additional operationalization of diffusion that we can test for with these data. This one is more subtle than reductions in variance for any one element. At the firm or report level, the choices on each discrete element should be more correlated if disclosures are interdependent. The distributions of a policy choice B conditional on each option from policy choice A will be more different from each other in Group-Two than in Group-One. If we know that a firm chose X for policy choice A it will affect their likelihood of having also selected Y for policy choice B when this actor learns from others. The same sources which prompted them to choose X should also increase the chances that they choose Y. This
conditional variation should be greater in the treatment than it is in the baseline case. An analogy from a very different setting might help make this operationalization clearer. Assume that an influential Hollywood star that shapes young males’ fashion choices is photographed wearing extra long shorts and tall wool socks instead of pants on a cool summer evening. In the ensuing weeks we might expect to find a higher prevalence of long shorts and a higher prevalence of tall wool socks in a random sample of young males. The influential star has made both more popular. We should also expect to find a number of teens who wear long shorts with tall socks and another group that wears neither. The distribution of the “tall wool socks” choice conditional on wearing “extra long shorts” will be significantly different than it will be conditional on “no extra long socks.” If one got the idea that tall wool socks were a good idea they also likely got the idea that extra long shorts were a good idea. Thus, tall socks should predict long shorts, and vice versa, more than they would have before this picture appeared.

5 Word Count, Readability and Other “Policies”

One key trait of any document, including a CD&A, is its length. While word count is a crude measure and blunt proxy, it is still an important variable. Similar to students facing a paper assignment without page length guidance, those who drafted CD&As had to decide how long to make the reports. Moreover, in this case, seemingly superficial choices such as length are actually more than stylistic. Since the regulation’s goal was clear disclosure, these choices about how to present information, and how much to present, are substantive. The length of these reports is likely correlated with other matters such as how detailed the information is, how much complex verbiage they use, and how much non-mandatory content is included. Additionally, the SEC’s own public comments suggest that length is an important dimension of these reports. It was one of the first things they commented on and looked at when determining how well companies were meeting the requirements. Report length was a tricky variable for compliance because the optimal length from the SEC’s perspective is a goldilocks policy. Some reports were too short and thus likely left out material information while others were far too long and obscured the important information
with unnecessary content and verbiage. In addition to word count, the SEC looked at the complexity of the language. They referenced readability scores which estimate the clarity of the reports, and their accessibility to the layperson. This section investigates whether there is evidence of learning and policy diffusion using these two metrics. In this context, public companies made “policy choices” about the length and language in their disclosures.

Visual inspection of the data in figure 1 is strongly suggestive of learning by the second group. It appears that those in the second group at least inferred what a reasonable length was and converged toward it. This figure displays the Word count distributions using a scatter plot and histograms. The scatter plot shows the Group-One firms on the left with March disclosures and then the gap before Group-Two data. The CD&As got notably shorter, by 20% or so. Importantly, this results differs from the audit committee reports which, as we saw earlier, were the same length in the two groups. More importantly, the distribution of word counts got substantially tighter in the second group. The first group shows that, given a vague and “principles based” directive, behavior varied wildly. The word counts are scattered, with good coverage, from about 1,500 to 15,000 words without an obvious and dense central tendency. Clearly the early adopters had no shared idea of how much to include and write. This is especially clear in the histogram.

The word counts in the second group become much more concentrated and the reports got shorter on average. They get more concentrated in two ways. First, the extreme outliers, especially at the top end, disappear. Second, the more moderate totals become much denser as well. The Group-Two histogram, for example, shows a much clearer central tendency than Group-One’s. The top half of table 1 evinces this same pattern in summary statistics. It shows that the mean and median length of the reports fell considerably, and that these differences were highly significant. More importantly, it shows that the standard deviation fell by even more and that this change was also very unlikely the result of chance. The table also reports various ranges in the two distributions which are wholly consistent with the figures’ visual story. The length of reports which were written with the benefit of observing others converged relative to the wide variety of report lengths that firms acting independently produced. These distributional changes are measures of the magnitude of the causal effect of having the opportunity to observe others. In the word count case, this opportunity led to
Figure 1: Plot and histogram of word counts for first and second group of policy adopters. The scatter plot shows word counts by proxy date including interquartile range and middle quintile range for the first and second groups. The histogram shows the same distributions. The word counts of CD&As varied wildly in the first group. They got shorter as a whole and less dispersed in the second group where the high outliers disappeared and the distribution was denser around the central tendency.

The dates in the second group have been altered to make the x-axis span similar to that in the first group for easier visual comparison. They actually span from early May to early July. The interquartile and 3rd quintile ranges apply to the entire group. Word count is measured from the beginning of the CD&A section to the beginning of the summary compensation table in all but a few idiosyncratic cases.

about a 20% decrease in the standard deviation of policies, and about a 15% decrease in the width of the middle third of policies for example. This finding is especially strong since the gap between the Group-Two and Group-One firms’ disclosures was only two months. This was barely even enough time for Group-Two firms to incorporate lessons from Group-One and yet we see strong convergence almost immediately.

In contrast, the same inspection and analysis of the Readability score data does not find much. (Summary statistics are in the bottom half of table 1). The two distributions look quite similar. The only notable difference is that the mean increases a little in the second group but this change is marginally significant. The standard deviation does decrease but by a small amount and this change is not significant. Little is evident in the plot and histogram (not shown). Unlike with the word counts, firms did not appear to converge in the second period. The lack of similar findings in readability scores is not surprising. While strong convergence would perhaps be the strongest evidence in support of the diffusion claim, the lack of convergence is in some ways supportive for two different reasons. Word count is more observable to the analyst, and more importantly, to the drafter. It is fairly easy to say “I’m
Table 1: Summary statistics comparing the word count and readability score distributions in Group-One and Group-Two. The CD&As got shorter and this length varied less in Group-Two. This is evident in the sharp decrease in the standard deviation and in the narrowing of all of the percentile ranges. The readability data tell a different story. There was essentially no difference between the two groups.

<table>
<thead>
<tr>
<th></th>
<th>Group-One</th>
<th>Group-Two</th>
<th>Diff (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word Count</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>5433</td>
<td>4230</td>
<td>1203 (.00***</td>
</tr>
<tr>
<td>Median</td>
<td>5213</td>
<td>4024</td>
<td>1189 (.00***</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>2519</td>
<td>2028</td>
<td>491 (.02**</td>
</tr>
<tr>
<td>Min–Max (Range)</td>
<td>664–14974</td>
<td>897–10505</td>
<td>4702</td>
</tr>
<tr>
<td>10%–90% (Range)</td>
<td>2444–8441</td>
<td>1828–6677</td>
<td>1148</td>
</tr>
<tr>
<td>25%–75% (Range)</td>
<td>3668–6590</td>
<td>2769–5321</td>
<td>370</td>
</tr>
<tr>
<td>33%–67% (Range)</td>
<td>4064–6109</td>
<td>3135–4680</td>
<td>500</td>
</tr>
<tr>
<td>45%–55% (Range)</td>
<td>4898–5580</td>
<td>3813–4163</td>
<td>278</td>
</tr>
<tr>
<td><strong>Readability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>15.30</td>
<td>15.61</td>
<td>-.32 (.10*)</td>
</tr>
<tr>
<td>Median</td>
<td>15.36</td>
<td>15.6</td>
<td>-.24 (.45</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.61</td>
<td>1.50</td>
<td>.11 (.43</td>
</tr>
<tr>
<td>Min–Max (Range)</td>
<td>10.2–19.7</td>
<td>12.2–19.2</td>
<td>2.5</td>
</tr>
<tr>
<td>10%–90% (Range)</td>
<td>13.0–17.16</td>
<td>13.7–17.9</td>
<td>.1</td>
</tr>
<tr>
<td>25%–75% (Range)</td>
<td>14.4–16.3</td>
<td>14.7–16.6</td>
<td>.0</td>
</tr>
<tr>
<td>33%–67% (Range)</td>
<td>14.7–16.0</td>
<td>14.9–16.2</td>
<td>.0</td>
</tr>
<tr>
<td>45%–55% (Range)</td>
<td>15.2–15.6</td>
<td>15.3–15.7</td>
<td>.0</td>
</tr>
</tbody>
</table>

The numbers in parentheses are the width of the percentile ranges. The “Diff” column for the range variables refers to the difference between the ranges.

not sure how long to make this” and quickly get a sense of how others have dealt with that uncertainty. It is more difficult to get a sense of language complexity and to tailor a report to meet some standard. We should, and do, see more convergence where learning lessons from others’ work is easier to do. Secondly, as I discuss in depth below, the fact that the second group’s readability scores did not decrease or converge suggests that the patterns we saw in the word count data were not merely a response to the SEC giving early feedback that the first wave of CD&As were too long and technical.

5.1 Alternative Explanations for the Word Count Data

There are a couple of very plausible alternative explanations for the differences we have noted which have nothing to do with learning or policy diffusion. The first of these two explanations is that the difference in the two groups’ disclosures results from differences in
the groups’ composition. The second alternative explanation is that the second group was responding to additional information provided by the SEC that was unavailable to the first group.

As we noted earlier, while the “treatment” was “assigned” exogenously, the two groups are not perfectly balanced and there are some differences which could cause spurious diffusion effects. The two most plausible and observable compositional differences are 1) different distributions of compensation Consultants which often help the firms write the reports, and 2) different distributions of Industries. These differences could affect the mean and/or variance of the word count distributions. They could affect the mean if, for example, financial firms which were heavily concentrated in Group-One had more complex compensation schemes that required lengthier disclose, or if the Watson Wyatt compensation consulting firm, which serves many of the retailers that had later proxies, tended to advise clients to include less information. These compositional factors could affect the variance if two firms in the same industry, or two firms that use the same consultant, likely have more similar reports than two firms chosen at random AND if the second group is dominated by one or two consultancies or industries reducing the variance.

Figure 2 contains quite a bit of information on multiple figures but the takeaway conclusion is pretty straightforward. Together, these graphics suggest that these alternative explanations are not supported at all and do not undermine the diffusion claims. The top half of the figure compares the average word count for each consultancy and industry by group. It shows that while there are some consultancies and industries that tend to produce longer reports than others, these differences are relatively mild. More importantly, it shows that compositional factors were not the primary cause of the shorter reports in the second group. Each industry’s and consultancy’s reports changed in the second group. The second group did not change because the distribution of industries and consultancies in it changed. The lower panels refute the possibility that the reduced variance was the result of a small number of industries or firms, presumably with correlated word counts, dominating the second group. The figures show the distribution of industries and consulting firms in each of the two groups starting with most common. One large category on the left side would indicate a high concentration of reports from one consulting firm or industry. Consultancies
and industries were more evenly distributed in the second group which should, all else equal, increase the variance in it.\footnote{This is especially true since the second most common consultant category in the second group was “none.”}

While it appears that group composition cannot explain the differences in the word count distributions, it is likely that the SEC providing initial commentary before the second group acted had some effect. At the same time, it is unlikely that this intervention explains all of the change or undercuts the apparent policy diffusion. Unfortunately, the SEC interfered with our natural experiment. They did so by, of all things, looking at word counts and readability scores after the first hundred or so CD\&As were published.\footnote{They cited almost identical means to those I calculated - Christopher Cox 3/23/07 remarks at USC School of Business.} Shortly after the first wave of public disclosures, SEC chairman Christopher Cox began citing these data in public speeches while airing the commission’s grievances with the initial batch of disclosures.

In the very first public comments, in a speech on March, 8 2007, Cox complained of seeing 30-40 page long disclosures written in legalese. In a speech a couple of weeks later, he expanded on this theme by complaining about the length and complexity of the initial CD\&A reports.\footnote{“Already we’re seeing examples of over-lawyering that are leading to 30-40 page long executive compensations in proxy statements...I have to report that we are disappointed with the lack of clarity in much of the narrative disclosure that’s been filed with the SEC so far.” (3/23/07)}

Given these public comments it would be surprising if some of the changes we observed were not attributable to the SEC’s initial reactions. On the other hand, it also appears unlikely that they explain all of the changes. While the SEC discussed section length, Chairman Cox spent much more time discussing readability. He complained of legalese, explained the Flesch-Kincaid and Gunning Fog readability metrics, and noted that some reports were harder to read than Ph.D. dissertations.\footnote{In the March 23rd speech he said “the good news is, Senator Barack Obama and the SEC’s General Counsel, Brian Cartwright, could read and understand these disclosures. That’s because they were both President of the Harvard Law Review. And a 34.86 (Flesch...
that while the SEC had problems with the length of some of these reports, the primary issue was readability. Firms in the second wave clearly did not heed these comments regarding readability. The readability scores indicate that the reports got slightly harder to read. This fact at least suggests that firms did not react strongly to the Chairman’s comments and thus, that the comments did not singularly cause the word count changes either. Relatedly, these were only initial comments, not an official policy statement. The chairman explicitly said that the SEC was “giving people a grace period,” and that they would enforce the plain English requirements “increasingly strictly” in the coming year. Additionally, the word count comments were limited to citing the length of the very longest reports which may explain the lack of 15,000 word outliers in the second group but is a less convincing explanation for the overall convergence. Finally, the SEC continued mentioning word counts and readability throughout the spring.\(^\text{10}\) As the SEC and press continued publicizing these problems, we might expect to see these comments affect later proxies in increasing amounts. This did not happen. The mean and median length of the 50 earliest proxies in Group-Two (early May) were the same as those from later May and June.

5.2 Other Choices: Tables, How to Start, and Disclosing Peers

So far we have observed convergence to shorter CD&A sections in the second group of responders. This is suggestive of learning from others. This finding was robust to plausible alternative explanations. We also looked at the readability scores, but did not observe the same convergence. Length and readability were the two metrics that the SEC first looked at in evaluating CD&As itself. Investigating a couple of other policies provides more evidence for the diffusion claim. The three other policy variables I coded are: 1) how many non-mandatory tables and figures a firm included in its disclosure, 2) what the firm discussed first in its disclosure, and 3) whether the firm disclosed a list of peer companies to which

\[^{10}\text{For example, John White, the Chairman of the SEC’s Corporation Finance division which oversees the disclosures, discussed length and plain English standards in a May 3rd speech.}\]
Figure 2: Summary of word counts by consultancy/industry and group and distribution of consultancies/industries by group. The top half of the figure shows that CD&As got shorter across the board in the second group. The latter adopters’ shorter reports were not the result of having a disproportionate number of firms in industries that tended to have shorter reports or that used consultancies that tended to produce shorter reports. The bottom half of the figure shows that the reduction in variance in the second group is not the result of having a less diverse composition of industries or consultancies represented in it. In fact, the data points in the second group came from a less concentrated set of industries and consultancies. Together, the figures strongly refute the very plausible alternative explanations that group composition drove the results.

**Compensation Consultant**

**Industry**

**Word Count by Group**

**Concentration by Group**

Large dots indicate global means for each consultant and industry irrespective of Group. Consultancy information collected for those firms for which it was readily available in the Proxy Report. The NA group is likely smaller in reality. Industry data based on two digit SIC codes.
it benchmarks pay. All three were discretionary, though the SEC suggested the third as an option early on.

Importantly, all three were not included in the SEC’s initial public feedback and thus avoid those potential confounds. Moreover, like length and readability, choices about tables and including a list of compensation peers (but not the first section of the report) are more than merely stylistic in this context. Both are closely connected to the regulations’ goals of increased transparency, disclosure, and clarity. They are independent of how much a firm pays its executives but pieces of how a firm explains its executive pay. Cogently explaining and disclosing information, in good tables instead of lawyerly prose for example, was a substantive aim of the regulations which required stylistic choices as part of compliance.

We have all probably looked at someone else’s work when we are unsure how to start or unsure about what level of detail to provide. We are looking for the same thing when analyzing both of the number of tables and the first section “policies.” In both cases, we are looking for convergence to one or two options in the second group relative to a more uniform distribution in the first. The number of tables variable is actually a count of the number of non-mandatory tables, graphics, and figures. I counted anything that was not paragraph text or a one column bulleted list as a “table.” The “first section” data represent the first bold or italicized section in the CD&A piece of a proxy statement. For most academic papers, this variable would be coded “Introduction.” I first collected the exact wording of the first section header and then eliminated small variations to get a more manageable, workable, and substantively useful “first section data” variable. The majority of these data points still represent the exact wording, (nearly all of the “overview” data points come from documents where the exact word “overview” denoted the first section) but some represent small changes or subjective judgments and distinctions.

The convergence we are looking for in the Number of Tables policy choice is evident in the comparison of the distributions (figure 3). More firms decided not to include any voluntary tables or figures in the second group. While zero tables was the most popular option in both groups, it comprised only a small plurality of data points, 28%, in the first group and

11As noted in the word count section, I focused on the body of the CD&A disclosure stopping before the mandatory “summary compensation” table.
jumped to 42% in the second. As the figure also shows, the first group featured a small but substantial number of CD&As (14%) that utilized at least five tables. The second group only included a few (6%) with five or more. A key summary statistic for the differences in the distributions is the dispersion, the variance divided by the mean. The first group is more over dispersed (ratio larger than one), that is, less concentrated, than the second. The variance is three times the mean in the first group (7.3/2.4) while it is only 2.25 the mean (3.4/1.5) in the second. Additionally, after creating a “five or more” variable, a chi squared test on the two distributions (10.57 on five degrees of freedom) shows that they are roughly close to significantly different (p=.06) at the 95% level. Moreover, this is a two way test and the distributions are different in the way we would hypothesize them to be. In short, the numbers of voluntary tables in the two groups are both different and change in the ways we would expect them to if those in the second group were learning from those in the first.

We can also see this convergence statistically in the models summarized in table 2. There are two ways to estimate this convergence. One is to model the number of tables as a count variable and investigate whether the number of tables decreases significantly in the second group. This approach measures the distribution’s convergence to the left and captures both the shift into the zero option and the movement away from reports with a large number of tables. The other way to model this convergence is to treat the decision to not include any tables as a dichotomous one and simply focus on this choice. The table reports the results of four such models: two negative binomial models of the number of tables,\(^\text{12}\) and two probit models of the choice to not include any tables. One of each of the two types of models does not include any controls. The second model in each group controls for compensation consultant and industry effects with indicator variables. The models show that either conceptualization of convergence is strongly supported by the data. The counts decreased and the probability of not including any tables increased by substantial and statistically significant amounts in the second group. There are a number of ways to substantively interpret these models. For example, as the row near the bottom of the tables

\(^{12}\)Because the count data was somewhat over-dispersed (variance = three times the mean) I report the negative binomial model, but the results were very similar assuming a Poisson distribution.
Firms varied widely in the number of tables and figures they used. Many used none or one while a few used more than ten. More importantly, as with the word counts, firms converged toward a common practice in the second group. In the first group, zero was the most popular option with about 28%. In contrast, in the second group, firms converged so that over 40% used zero tables.

"Tables" includes everything (graphics, tables, figures, two column lists) that was not text in either paragraph form or one column bulleted lists.
Table 2: Count models of number of tables and Probit models of having zero tables. The table summarizes two negative-binomial models for the number of tables and two probit models for “zero tables” option. Both models with and without controls for consultancy and industry effects, show a consistent, substantial and statistically significant effect of being in the second group. Firms that had the opportunity to observe the first group’s disclosures used fewer tables and were more likely, by 10-14%, to not use any at all.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Neg. Binomial: Table Count</th>
<th>Probit: Zero Tables?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
</tr>
<tr>
<td><strong>Main Effect</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group-Two</td>
<td>-.48</td>
<td>.16</td>
</tr>
<tr>
<td>Constant</td>
<td>.89</td>
<td>.11</td>
</tr>
<tr>
<td><strong>Consultant Effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cook</td>
<td>-.43</td>
<td>.36</td>
</tr>
<tr>
<td>Hewitt</td>
<td>.33</td>
<td>.29</td>
</tr>
<tr>
<td>Mercer</td>
<td>.79</td>
<td>.25</td>
</tr>
<tr>
<td>Towers</td>
<td>.29</td>
<td>.27</td>
</tr>
<tr>
<td>Other</td>
<td>.38</td>
<td>.21</td>
</tr>
<tr>
<td>None</td>
<td>-.35</td>
<td>.37</td>
</tr>
<tr>
<td><strong>Industry Effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manufacture</td>
<td>-.61</td>
<td>.21</td>
</tr>
<tr>
<td>Trans/Comm</td>
<td>-.58</td>
<td>.36</td>
</tr>
<tr>
<td>Retail</td>
<td>-.66</td>
<td>.36</td>
</tr>
<tr>
<td>Finance</td>
<td>-.41</td>
<td>.32</td>
</tr>
<tr>
<td>Service</td>
<td>-.54</td>
<td>.35</td>
</tr>
</tbody>
</table>

Increase in probability of zero tables

- **11%** (4%, 18%)
- **10%** (1%, 18%)
- **14%** (2%, 26%)
- **14%** (0%, 28%)

<table>
<thead>
<tr>
<th>N</th>
<th>L-Likeli, $\chi^2$(df), P</th>
<th>238</th>
<th>238</th>
<th>238</th>
<th>238</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-446, 9.31(1), .002</td>
<td>-436, 29.7(12), .003</td>
<td>-151, 5.37(1), .02</td>
<td>-140, 27.2(12), .007</td>
<td></td>
</tr>
</tbody>
</table>

Predicted probabilities calculated using delta method (SPost in STATA 10.1 (Long and Freese 2005))

show, Group-Two firms were 10-14% more likely to not include any voluntary tables. These differences were statistically significant and are not driven by consultancy and industry effects.

Similar to the number of tables, the way that a firm starts its CD&A is a potential indicator of learning. I coded for the *First Section* in a firm’s CD&A as described above. This variable is both the least substantive and most problematic to measure and thus I only mention it briefly. Nevertheless, it fits the diffusion pattern. These data are summarized in figure 4. As trivial as it seems, the “policy” of beginning with “overview,” appears to have diffused. The visual story is similar to that we observed when counting tables. The distribution of first sections was relatively uniform in the first group with the most popular option, “objectives,” comprising about 18%, and the second most popular, “overview,” about 17%. In contrast, firms in the second group converged to starting with “overview.”
Figure 4: Histogram of the substance of the first section header in each CD&A. The behavioral pattern this table captures is similar to that we observed when using the tables policy. While there was not one standard way to begin a CD&A report, one option, starting “overview” became much more popular in the second group. As before, this convergence to one option, which appeared to be roughly as popular as a few other alternatives in the first group, suggests learning and diffusion.

Most of the first section codings were literally word for word. A few reflected the author’s judgment to collapse permutations relating to the same substance into a manageable number of categories. Eight percent of firms used it compared to about 15% for the second most popular option “objectives.” These differences are fairly large. There are so many ways a firm can start its report and this choice is so inconsequential that any amount of convergence is a strong indicator of learning. While we are limited in our ability to run statistical tests for variance reduction with this nominal variable, in part because doing so requires tautologically selecting on the choice that became popular, we can estimate regressions (not reported) to assess the apparent increase in the popularity of the “overview” option. The “second group” variable is a significant predictor, by 9-13%, of starting with “overview” with the same controls as earlier.
5.3 Disclosing Compensation Peers

The fifth and final element that I coded was whether a firm disclosed, by name, the other firms that it indexes its compensation to in a Peer List.\footnote{This coding scheme, where I simply looked for a list of peers, assumes that firms that actually do not index to a set list of peers were evenly distributed across the first two groups since they are lumped in with those that index, but did not disclose a list.} I also coded whether it disclosed this information in a table or in the text. So far we have seen convergence in length, use of tables, and decisions about how to begin. This fifth policy does not offer much support on its own. Shortly we will see that it does offer strong support for the diffusion claims when combined with the other policies. This peer disclosure was another element of the reporting that the SEC suggested, but which was not required. It was dichotomous and thus it is less clear what convergence would look like.

Strong evidence of convergence in a dichotomous choice would be a shift closer to either all “yeses” or all “nos” in the second group. This appears not to have happened. In the first group, 57% of firms disclosed their pay peers by name while in the second group 52% did. Not surprisingly, these binomial distributions are not significantly different from each other. Moreover, of those that did disclose a list or peers, approximately 60% of them did so in tables in both the first and second groups with the rest doing so in the text.\footnote{This later fact offers a fragment of diffusion evidence since we already noted a decrease in the use of tables generally in the second group but the use of a peer table remained at its Group-One level suggesting a relative increase in its popularity.} On the one hand, it would be ambitious to expect to find diffusion in every element given the relatively short time frame between the two groups and the fact that they do have the capacity to make these decisions independently. Admittedly, on the other hand, this is the type of decision in which we might expect to see increasing conformity to common practices.
6 A Different Measure of Diffusion: Conditional Variation Within Reports

So far we have looked at five policy decisions a firm had to make when crafting their CD&A for evidence of convergence. If policies are diffusing, we should see less variation in a given policy’s distribution. We should also see interdependence between policies within a firm’s report in the second group. The common practices should not only appear more frequently in the second group, but they should appear together or not appear together more frequently at the report level.

This expectation is elaborated in detail earlier using fashion trends as an analogy. Another way to think about this conceptualization of diffusion is to think about an experiment to investigate whether students illicitly work together on take home exams. Imagine giving one section of a class the exam in class and the other (equally capable) section the same exam as a take home. Observing better and less varied scores in the take-home exam section would point to cheating just as the metrics above point to diffusion. Nevertheless, the real test is whether the answers in the take-home treatment are more correlated. Is getting question number one right a stronger predictor of getting question number two wrong in the take home case than it is in the in-class exam treatment for example?

This section investigates diffusion by looking for these types of correlated choices. If some practices are diffusing as firms look at others’ reports, we should see some of these practices converge together in the same second group reports. For example, we are looking for differences in the distributions of the “include peer list” conditional on the “begin with overview” decision. The distributions of “peer list” conditional on “overview” and “no overview” should be more similar to each other in the first group than in the second. A choice on one part should predict a choice on another more in the second group than in the baseline case.

Some elements will be almost mechanically correlated (e.g. word count and the number of tables) so we cannot analyze them, but others will not. There are not too many reasons other than learning and interdependence to see “overview” and “zero tables” adopted together at rates that far exceed the independent base rates in the second group. This section conducts
this investigation using three dichotomous choices: 1) including any voluntary tables, 2) including a list of peers, and 3) beginning with an “overview” section (the most popular option). It finds strong evidence to support the third diffusion hypothesis. These three choices were almost perfectly independent in the first group but significantly interdependent in the second. This is in many ways the strongest evidence that firms in the second group followed models from those in the first.

Table 3 contains information about choice correlation in the second group. Because many of those that included a list of peers included them in a table, I modified the “no tables” variable for this analysis. It now applies to tables other than peer lists. A firm that gets a “1” for “no tables” either did not include any tables, or only included one table in which it reports a peer list. The top part of the table compares the expected incidences of pairings of choices under the assumption that they were independent with the actual observed incidences. For example, the upper left cell compares the expected rate of reports that start with “overview” and do not include tables to the observed rate. The predicted or null hypothesis (independence) rate is just the product of the two unconditional base rates. For example, in Group-One 34% of reports didn’t include any tables and 18% started with “overview.” Thus, the predicted rate, assuming independence, is (.18 * .34 = .06). In this case, as the upper left cell of the table shows, the predicted and actual rates hardly differ at all.

As the Group-One section of the table shows, the observed incidences were almost exactly those which we would have expected if the three choices were independent. In the first group, choices about voluntary tables, disclosing compensation peers, and starting with “overview” did not move together at all. The distribution of any one of these choices, conditional on another, were virtually identical. In contrast, the Group-Two panel suggests some deviations from independence. Not including any tables and not disclosing peers appeared together more frequently in Group-Two reports than they should have if the choices were independent. Similarly, not disclosing peers and starting with “overview” appeared less frequently together than we would expect.

The bottom half of the table investigates these relationships more rigorously. It summarizes two probit regressions, one for Group-One and one for Group-Two, with the “disclose
peers” variable on the left hand side and “no tables,” “start with overview,” the interaction between them, and controls for consultant and industry effects on the right hand side. Here, I am explicitly testing for correlation. I am not making causal claims but merely using the method to estimate interdependence with standard errors and controls. I report the models with “disclose peers” on the left hand side to demonstrate how the choices were linked in the second group, but I could have done the other permutations as well.

We should see relationships between the key choices in the Group-Two models only. Whichever we put on the right hand side should “predict” an outcome on the left hand side. This is exactly what we see. The Group-One models evince almost perfect independence. The Group-Two models tell a different story. Both “no tables” and “start with overview” decisions are substantially and significantly related to disclosing peers. Those who chose not to use tables and figures were also much more likely to not disclose compensation peers while those who started with “overview” were more likely to disclose their peers. The very bottom part of the table provides an idea of the magnitude of these connections. Among firms that did not start with “overview” (a substantial majority) those that did not use tables were 30 percentage points more likely to also not disclose peers. Those with tables and an “overview” start were 31 points more likely to disclose peers than those with tables and no “overview start.” Finally, those without tables and without “overview” starts were 61 points less likely to disclose peers than those with tables and overview starts.

These findings strongly point toward firms in the second group observing and learning from those in the first. It only took a couple of months for policy choices which were unrelated in the first group to move together in the second. Substantively, these findings also point to potential trouble for the SEC and for meaningful regulatory impact. Not only does it appear that these reports which were supposed to be customized and avoid “boiler plate” became more similar to each other very quickly, but it appears that a few different “types” started to emerge. One of these types was a low disclosure type. We see this in the strong relationship between not using tables and not disclosing compensation peers. All else equal, both of these traits would make a report less informative by reducing the amount of information available and/or the accessibility of the information that was provided. It appears that by May and June in the first year of the regulations, a blueprint
Table 3: Choice independence in the first group and choice correlation in the second. The top part of the table compares the observed combinations of three binary choices to the predicted values if the choices were independent. The same choices appear to be almost perfectly independent in the first group and appear to be interdependent, correlated, in the second. The bottom part of the table summarizes probit models to test for this interdependence. There was absolutely no relationship between the three decisions in the first group, but they moved strongly and significantly in the second. That is, for example, in the first group, not using voluntary tables was not associated with an increased or decreased likelihood of disclosing compensation peers, but in the second group it was.

<table>
<thead>
<tr>
<th>Choice Combinations</th>
<th>Group-One</th>
<th>Group-Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted if Indep.</td>
<td>Actual</td>
<td>Difference</td>
</tr>
<tr>
<td>No Tables and Overview Start</td>
<td>6%</td>
<td>7%</td>
</tr>
<tr>
<td>No Tables and No Peer List</td>
<td>15%</td>
<td>15%</td>
</tr>
<tr>
<td>No Peer List and Overview Start</td>
<td>7%</td>
<td>6%</td>
</tr>
<tr>
<td>Base Rates</td>
<td>No Tables: 34%, Overview: 18%, No Peer List: 43%</td>
<td>No Tables: 48%, Overview: 27%, No Peer List: 48%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Probit: Include Peer List?</th>
<th>Est.</th>
<th>SE</th>
<th>P</th>
<th>Est.</th>
<th>SE</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Tables</td>
<td>.06</td>
<td>.30</td>
<td>.83</td>
<td>-.78***</td>
<td>.32</td>
<td>.01</td>
</tr>
<tr>
<td>Overview Start</td>
<td>.57</td>
<td>.42</td>
<td>.17</td>
<td>1.01**</td>
<td>.49</td>
<td>.04</td>
</tr>
<tr>
<td>No Tables x Overview Start</td>
<td>-.42</td>
<td>.69</td>
<td>.53</td>
<td>-.20</td>
<td>.64</td>
<td>.76</td>
</tr>
<tr>
<td>Consultant Controls...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry Controls...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Change in Probability of Disclosing Peers

Reports that do not start with “Overview:” Diff from Tables to No-Tables: - 30% (-7%, -53%)
Reports with Tables: Diff from No-Overview-Start to Overview-Start: 31% (7%, 55%)
Switch from Tables and Overview-Start to No-Tables and No-Overview Start: -61% (-38%, -83%)

Predicted probabilities calculated using delta method (SPost in STATA 10.1 (Long and Freese 2005)) 95% CIs in ( )

for less informative CD&As was established and spreading. Of course, this also means that a blueprint for a more informative CD&As was doing the same. Nevertheless, a substantial number of firms appear to have looked at other reports and learned how to produce a CD&A which provides relatively little information.

7 Substantive Impact: Annual Performance Bonuses

Thus far we have seen that companies that got to comply with the disclosure requirements later appear to have learned from those that went before them. This learning, and diffusion of compliance policies, is evident in metrics such as the length of reports and their use of
tables. Perhaps even more importantly, ostensibly discrete choices, e.g. whether to disclose peers and whether to use tables and figures, predict each other in the second group but not the first. In this case, since the regulations concerned conveying information, these seemingly formal, stylistic, and almost trivial variables are substantively significant in their own right. Nevertheless, ultimately we care about the quality of the disclosure of important elements of executive compensation. In this section I analyze the quality of disclosure by focusing on discussions of annual cash performance bonuses and find less forthcoming disclosure in the later adopters. While the exact mechanism is hard to untangle, it appears that later adopters either learned how to disclose less, or inferred what limited, but respectable, compliance looked like from the first group.

The regulations’ outcome goal was clear and thorough disclosure and explanation. In this section, I investigate the link between ostensible learning and diffusion in more stylistic matters and substantive disclosure of actual pay practices. I do this by focusing on firms’ explanations of their annual performance bonuses. These annual cash bonuses, along with salary and long term equity pay, are an important part of nearly every companies’ executive compensation program. I focused on this element of compensation because every company will have some sort of process for determining annual bonuses. The processes and procedures are the types of things that the regulations sought to bring into the open. Normally they comprise a set of key performance metrics and targets which then map into compensation amounts. Companies would have had a lot of flexibility and uncertainty about how much detail they would disclose and how much they would explain to the public.

This analysis shows a consistent pattern of less disclosure among the second set of firms across four distinct measures of their annual bonus reporting. While the process and mechanisms are unclear at this point, the data strongly show that firms in Group-Two, which got to observe the first group’s disclosures, learned to disclose less about annual bonuses. More of them shared less, less detailed, and less clear information than Group-One companies did when acting independently.

I coded four elements of these disclosures of annual performance bonuses. As before, I focused on explanation and disclosure rather than dollar amounts. I tried to capture how well companies explained and justified their annual bonuses, not how warranted the bonuses
actually were. A company detailing that a $.01 increase in sales would lead to a $10,000,000 bonus would do better in this coding than a company that did not explain its bonus process even if it produced much smaller payouts. To do this, I broke the discussion of annual cash bonuses into four elements which I then coded with the help of research assistants.\textsuperscript{15} We broke the reports into a small number (between two and four) of performance categories for each element after looking over a few reports. These categories roughly approximate qualitatively distinct “types.” Lower numbers are always indicative of less disclosure but the scales are unique to each of the four elements and are not directly comparable. These four elements, and a sense of the guidelines we used for coding each, are as follows.\textsuperscript{16}

1. \textit{Process}: How well did the report explain, in the abstract, the process and inputs which determine annual bonuses? In practice, this process description normally includes things such as the metrics on which executives’ performance is measured (e.g. earnings per share, return on equity, sales growth), predetermined targets for the relevant metrics (e.g. $2.02 per share was our “threshold” target), and a mapping from performance to bonuses (e.g. “achieving the threshold target would result in a bonus of $X for the CEO). The coding scheme had four categories for “process.” Zeros were reports that said essentially nothing about the process for determining bonuses. Ones were given to reports that mentioned, for example, the metrics they use with little other detail (e.g. “our bonuses are based on earnings per share...”). A Two was given to a report that described metrics and targets in more detail. Finally, reports that described these elements in great detail and beyond such that one could largely recreate the process based on reading the report were coded as Threes. Category three reports also often described how the metrics, targets and/or other inputs varied for each executive and did other things which conveyed exemplary precision.\textsuperscript{17}

\textsuperscript{15}The reports were coded in random order and irrespective of “group.”
\textsuperscript{16}While we had guidelines for each element, and certain things we were looking for to distinguish a category one from a category two for example, we were ultimately being cognizant of how much information each report conveyed to a educated but non expert reader.
\textsuperscript{17}We did not “punish” companies for not having thorough processes. If a report was clear that its process did not include these elements but did convey how things worked it would
2. **Justification:** Did the report justify and explain the rationale behind the elements in the process? For example, did they give a reason for using the financial metrics they use (“we believe earnings per share is the best performance metric because...”) and/or explain where the performance targets came from? (“our threshold target assumes 5% growth over last year”) This variable was coded Zero or One. A One was given to a report that provided even minimal justification.

3. **Implementation:** This variable measured how well the report explained how the abstract process was applied to the most recent year to produce a dollar amount. Some reports simply pointed to the annual bonus number in their “summary compensation table” (or said nothing at all), and these reports got a zero for implementation. Others said, for example, how the company did on the relevant performance metrics (from the process) in FY 2006. This level of detail was given a one. Other companies went further and basically walked through the process with the past year’s data, either in prose or a table, going from targets, to performance, to the mapping to cash bonuses. These reports were coded Two.

4. **General Clarity:** The fourth element measured how well the discussion of annual bonuses met the “plain English” goal. This variable is simply the coders’ (talented, but non-specialist, undergraduates) impression of how clear, comprehensible, and free of legalese and jargon the discussion was. This variable did not take into account the amount of information provided. It simply measured the clarity of what was written and thus it was often difficult to provide a lot of detail about process and do well on this clarity measure. Reports were coded from zero (least clear) to two (very clear).

The outcome of interest is pretty straightforward. We want to see whether there were systematic differences in the level of disclosure between the two groups. In other words, did the fact that Group-Two firms appear to have learned from their predecessors affect how much information they provided about annual bonuses? Because of the quasi-experimental design, we should not see differences unless the second group learned from the first. It turns out that we see substantial differences. Across all four metrics, companies in the second receive a high score.
group were less forthcoming about their annual bonuses. They provides less information about, and justification for, their processes. They offered less detail about where the 2006 executive bonuses came from, and the information they did provide was less clear than that in the earlier disclosures.

The differences in disclosure are apparent in the distributions of all four of the annual bonus variables summarized in figure 5. We see a shift from “good” disclosure to “bad” disclosure from Group-One to Group-Two across all four components. Group-One companies provided the “best” process explanations more often (by 13 percentage points) while Group-Two companies provided minimal disclosure more frequently. Group-Two companies had more category zero process disclosures, no information, (8% to 4%) and more category one process disclosures, minimal information (30% to 18%). Similarly, they were slightly less likely to provide a justification for their processes. Group-Two companies were also about nine points more likely to say nothing about how their process was implemented in the past fiscal year, and eight points less likely to provide detailed implementation explanations. Finally, according to the coders’ impressions, there was about a ten point decrease in clarity, from the high clarity to the moderate clarity, from Group-Two to Group-One. This last finding may actually understate the real differences in “plain English” performance. As we have seen, the Group-One companies provided more information. All else equal it is harder to provide more information and provide information more clearly which works against Group-One firms. In sum, the magnitudes vary a little, but the pattern does not. Later compliers were consistently less detailed, clear, and forthcoming in their disclosures of their short term performance bonuses than those that complied independently in the first group.

Two of these four apparent differences are highly significant when controlling for consultancy and industry effects as before. (Details of these probit and ordered probit models are not reported.) The two groups’ disclosure of the processes behind their annual bonuses are very different as Group-Two firms disclosed much less information. The ordered probit coefficient on the Group-Two variable when controlling for consultants and industries is $-0.63$ (SE=.16, p=.000). Substantively, based on predicted probabilities derived from the models, a company in Group-One was about 21 percentage points more likely to fall into the “most process disclosure” category, while a company in Group-Two was 15 points more likely to
Figure 5: Distributions of four elements of annual bonus disclosures and explanations by group. The four panels compare the early (independent) and later (potential for diffusion) compliers. They show a pronounced shift toward less thorough, less detailed, and less clear disclosure. Firms that could observe the early adopters shared less information about their annual bonus processes, procedures, and inputs than they would have if they acted without seeing what others disclosed.

*Bonus Process*  
*Bonus Process Justification*

*Bonus Implementation*  
*Bonus Disclosure Clarity*

N=263 (144 in Group-One, 119 in Group-Two) for all four variables
fall into the “little process disclosure” category and six points more likely to fall into the “no process disclosure” category.

Similarly, the ordered probit coefficient on the Group-Two variable regarding discussions of bonus implementation was \(-.34\) (SE=.17, p=.04). A company in the second group was 11 points less likely to be in the “most implementation disclosure” group and 10 points more likely to be in the “no implementation disclosure” group. There was no evidence of an effect on justification and the apparent difference in clarity is not significant at conventional levels (p=.20), but it is in the same direction as the other differences. Nevertheless, the raw data and statistical models strongly suggest that being in Group-Two had a substantial, and negative effect on the amount and quality of disclosure. These companies were less likely to provide the important details such as the performance goals and criteria behind the annual bonuses they gave their executive officers.

In short, given the opportunity to observe others, public companies disclosed less information than they would have if they had to decide what to disclose independently in a vacuum. At this point it is not clear whether later compliers followed particular models which led them to disclose less, or if they were inclined to disclose less and saw enough minimal disclosures to feel comfortable doing so. Future work, including work using this design, can investigate these more nuanced mechanisms and decision making models.

8 Discussion and Conclusion

"When," “how,” and “how much,” private actors comply are absolutely fundamental issues of politics and public policy. More specifically, understanding the factors that shape the practical meaning and magnitude of compliance is essentially to understanding legal and regulatory implementation. It is thus crucial to understanding the interactions between public and private actors which ultimately determine policies’ impact on societal outcomes. Diffusion mechanisms, which are much more familiar and thoroughly studied in the policy creation context, play an important role in policy implementation as well. The SEC has fortuitously provided a natural experiment in a real and substantively important setting. The Compensation Discussion and Analysis requirements, particularly the fact that compliance
proceeded in quasi-random waves, offers unique leverage into all of these questions. This case allows us to cleanly observe and measure the effects of diffusion in implementation. It demonstrates that diffusion is an important mechanism in shaping the law’s impact. Given the opportunity to learn from others’ compliance, public companies converged to common practices and disclosed less information.

Above, we could see what policy choices looked like when they were made based on common external signals (e.g. the regulations), but were otherwise independent from each other. We could then compare this distribution to the distribution of policies when firms could see what others had done. Without focusing on particular mechanisms, this investigation provided strong evidence of policy interdependence, learning from others, and diffusion. As just about any policy diffusion mechanism would predict, policies were more similar to each other when firms could see what others had done than when they could not. The length of compensation disclosures, decisions about optional tables, and even how the reports started, become more similar from firm to firm when firms could see what the first wave of reports looked like. Moreover, in the second wave, having adopted any one of the choices that appeared to be diffusing was associated with adopting the others. Some firms’ reports appear to have adopted multiple common policies based on learning from others. This is especially strong evidence that at least some firms did not simply respond to the regulations and write their reports; but rather, that they took a close look at what others did and went from there. Additionally, the impact of diffusion extends beyond seemingly stylistic choices about word counts and tables. It appears that on average, later adopters learned to disclose less, or learned how to disclose less, than they would have if they acted independently.

This study’s unique case and research design should enable a number of other investigations of policy diffusion and regulatory implementation in the near future. Knowing the full set of plausible policy options from the baseline group can, for example, allow us to see which traits make a particular policy choice more or less likely to become common. These traits might include things like company size or prestige. Similarly, we now have a good idea of which firms in the second group adopted the diffusing policies. We can attempt to identify the traits that are more or less associated with looking to others’ policies. Additionally,
firms have now completed three years worth of CD&As. Thus, we can take a longer view and investigate how these policies have evolved and whether those options that appeared to be diffusing in the months after the regulations diffused in the years after the regulations. Among other things, we can see what the firms in the original baseline group did the second time around. Finally, we can exploit the same variation in compliance timing to ask whether having to justify executive compensation and perquisites led to restraint. In this case, we would be comparing firms that already had perquisites on the books before the regulations were announced to those that knew that they would have to justify their practices beginning in the 2007 fiscal year.

The findings above offer perhaps the cleanest documented evidence of policy diffusion and interdependence in general and in the policy implementation context. This case strongly suggests that legal and policy implementation, and legal and policy impact, will proceed via a series of interdependent choices. The practical meaning of the law depends on which practices become common through interdependent processes. These could be highly “compliant” and impactful policies but, as we saw, firms may also learn from each other how to do relatively little and/or how to evade the spirit of the regulations. This ostensible tendency toward diffusion and interdependence is particularly important in “principles based” regulations. It is also very relevant in many judicial impact settings as court decisions are also often vague, flexible, and unspecific about compliance details. The CD&A case is especially instructive because the SEC was self consciously and explicitly pushing back against diffusion and interdependence. In fact, they sought and called for independent and unique disclosures. The level of convergence that we observed in a very short period of time suggests they were not succeeding. Principles based regulations will become more and more salient in the coming weeks and months. They are part of discussions about regulating balance sheets, compensation incentives, securities ratings, and credit default swaps. This study suggests that diffusion processes and interdependent decision making will jeopardize even well intentioned regulations which seek to allow for differences in firm behavior and enable customized disclosure.
References


