Toxic Workers

Michael Housman
Evolv

Dylan Minor
Kellogg School of Management, Northwestern University

December, 2014

Abstract

While there has been much work on finding and developing top performers in the workplace, less attention has been focused on how to manage those workers who are harmful to organizational performance. In the most dramatic form, such workers can generate enormous regulatory and legal liabilities for the firm, harm its reputation, and invalidate its corporate social responsibility program. We explore a large dataset of over 50,000 workers across 11 different firms to document a variety of aspects of workers characteristics and circumstances that lead them to engage in actual "toxic" behavior. We also find that removing a toxic worker or converting him into an average worker enhances organizational performance to a much greater extent than replacing an average worker with a superstar worker.

Keywords: organizational design, misconduct, worker productivity, ethics, corporate social responsibility, superstars
1 Introduction

There is an abundance of work exploring how to find, develop, and incentivize top
performers to enhance organizational performance (Lazear and Oyer (2007) and Gib-
bons and Roberts (2013)). While this is important, it is also important to consider a
less studied area: avoiding workers who are harmful to an organization’s performance
(Banerjee et al. (2012)). These workers could simply be a bad fit, leading to pre-
mature termination and a costly search for and training of a new worker. However,
even worse is a worker who harms the organization itself, whom we refer to as a toxic
worker.\footnote{We use the term toxic worker as shorthand for workers who engage in toxic behavior. A worker’s likelihood of engaging in toxic behavior is assumed to be a function of both personal characteristics and work environment. Hence, we are not suggesting that toxic workers are a group of workers who are doomed to engage in toxic behavior solely as a function of their inherent, fixed nature.} This type of worker engages in sexual harassment, workplace violence, or
some other breach of important company policy or law. At worst, such workers can
cost a firm billions of dollars, as evidenced by the example of JP Morgan’s so-called
London Whale incident with Bruno Iksil.\footnote{See http://www.bloombergview.com/quicktake/the-london-whale. In this case, it was ultimately not Mr. Iksil himself who was charged (he cooperated with authorities), but rather his supervisor and junior trader.}

In addition to causing the firm and its reputation to incur a first-order cost
and liability, toxic workers are also problematic for those firms pursuing corporate
social responsibility. After all, a firm’s workers are the frontline manifestation of its
values and ethics. A "responsible" organization with workers who are, for example,
sexually harassing others is being inconsistently represented and its reputation will
be harmed. Recent evidence shows that more responsible organizations also become
better performing ones (Eccles et al. (2014) and Flammer (2014)), which suggests a
secondary cost to an unethical workforce.

The origins of worker misconduct are varied. There is consistent evidence that in-
centives can at least nudge workers to create adverse outcomes (e.g., see Oberholzer-
Gee and Wulf (2012), Larkin (2014), and Minor (2014)). There is also evidence that
a worker’s personal characteristics are important in determining his ethical behavior
(e.g., see Ford and Richardson (1994) and Loe et al. (2000)). Thus, one approach
to managing toxic workers—and the approach we focus on in this paper—is simply avoiding them. By exploring the actual conduct and characteristics of over 50,000 workers across time and different organizations, we identify several predictors of toxic workers: overconfidence, poor job fit, and self-proclaimed rule followers.

Much of the past work on worker ethics has been based on surveys, self-reports, and intention-based outcomes (Weaver and Trevino (1999) and Greenberg (2002)). Bertrand and Mullainathan (2001) suggest that the mixed results of this past work likely stem from the challenge of empirically examining subjective data. Our paper complements this important work by linking personal characteristics with objective conduct outcomes across a very large sample. Further, we find that these factors can be identified before hiring a worker.

In addition, we find evidence that a worker’s work environment (i.e., his particular location within a firm) contributes to the likelihood of him becoming a toxic worker. This complements the work of Pierce and Snyder (2014) who show that a worker’s firm has significant effects on her ethical conduct in the setting of automobile emissions testing. In addition to showing that this effect is also present in a broader setting, we are able to compare the importance of an individual’s characteristics and which individual characteristics matter in determining actual outcomes.

We also identify and measure second-order effects of toxic workers on an organization. In particular, we find that such workers increase the likelihood that other workers will also become toxic. In addition, toxic workers exhibit an above-average quantity level of performance. Thus, as in Gino and Ariely (2012), we find that there is a potential trade-off of employing an unethical person: on one dimension they excel and on another dimension they are corrupt. However, we also find that toxic workers produce subpar quality of work. We then compare toxic workers with "superstar" workers and find that removing the former has a greater organizational productivity impact than adding the latter.

An alternative approach to screening out toxic workers is to reform them. With resource constraints it might not be feasible for some, if not most, organizations to pursue this second path. However, we find that a worker’s environment could possibly be changed to reduce the chances of the worker becoming toxic in the first
place.

The paper is organized as follows. The next section develops a theoretical understanding of the problem of toxic workers and explores how we can identify their origins. Section three presents our main empirical results. Section four provides a discussion, and our final section concludes.

2 Theoretical Considerations: The Person and the Situation

In this section, we consider a simple theoretical setting to illustrate the link between theory and our identification strategy. We begin by assuming a simple world where all workers are the same and all environments are the same. That is, the person and the situation are always the same. This will serve as a baseline that will then be modified by allowing for different individuals and different situations. In this setting, all workers can engage in toxic behavior in a given period. Once they engage in such behavior, they are dubbed a toxic worker.\(^3\)

In particular, first assume that \(P\) represents the probability that a person will engage in some toxic behavior\(^4\) in a given period. In a worker's first period, she has a \(P\) chance of engaging in toxic behavior. This means that she has a \(1 - P\) chance of working in the next period, assuming a toxic worker is removed from the worker pool. Hence, the chance that a worker makes it beyond period \(t\) is \((1 - P)^t\), \(^5\) which we denote as the survival rate \(S(t)\). In contrast, the chance that a worker does not

\(^3\)In the general case, with different individuals and situations, it makes more sense to define a toxic workers as those who are more likely to engage in toxic behavior than those who or not. However, here all workers are toxic workers in that sense, so we use the term toxic worker to denote when such a worker engages in toxic behavior.

\(^4\)Here, and throughout when we use the term toxic behavior, we are assuming the kind of behavior that is observable (or its effects are observable).

\(^5\)Note that with this setup it is equivalent to assume workers have a constant propensity to engage in misconduct and finally do so in a given period with probability \(P\), after which they are removed from the worker pool, and to assume workers always engage in a constant level of misconduct and are finally caught and removed with probability \(P\). For this study, what we will actually observe is the removal of a worker (i.e., a termination) who has engaged in toxic behavior.
make it to period $t$ is $1 - (1 - P)^t$, which we denote as the failure rate $F(t)$. Recall that the hazard rate, where $f(t)$ is the density of $F(t)$, is then defined as

$$\begin{align*}
h(t) & := \frac{f(t)}{S(t)} \\
& = -\ln (1 - P)^t (1 - P)^t \\
& = -\ln (1 - P)^t.
\end{align*}$$

The hazard rate tells us the chance that a worker will be a toxic worker at time $t$, given that she has not yet been toxic up to time $t$. Not only is this an intuitive measure to consider, but there is a long, rich history of estimating hazard rates. In this simple setup of a constant chance of engaging in toxic behavior for all periods, for all people and all situations, the hazard rate would then simply be a linearly increasing function of time, since $-\ln (1 - P)^t = -t \ln (1 - P)$. Of course, in practice, we do not expect this to be true. In fact, we expect that the hazard rate is likely a very complex function. For example, even if all people and situations were the same, it could take more than one period for a worker to engage in toxic behavior: for example, perhaps it takes more than one period to learn about and take advantage of an opportunity. In this case, $P$ would increase over time. Alternatively, perhaps a person is more likely to engage in toxic behavior during the formative days and months at a new position; as time passes, she becomes better integrated and less likely to be toxic. Many other possibilities abound. Thus, it seems important not to assume some ex-ante relationship between time and the hazard rate. In our setting, we will refrain from making this assumption for our baseline hazard rate by instead specifying an overall hazard rate of engaging in toxicity for a particular person over time as

$$h(t) f(t|X),$$

where $h(t)$ is allowed to have an arbitrary relationship between the hazard rate and time and can be viewed as the average relationship between time and the hazard.
rate. In contrast, the function $f(t|X)$ takes on a value greater or less than 1 as a function of (potentially) both the person and situation across time. Specifically, $f(t|X)$ is a function of the person and the situation at time $t$, which are captured by the matrix $X$. In its simplest form, we could assume a world with workers that are sometimes toxic at a periodic rate of $h(t)$ and another set of workers who are never toxic. Thus, $f(t|X)$ would simply be an indicator function, taking on the value of 1 or 0, depending on whether the worker is toxic.

For a richer example, assume after 365 days an average remaining worker has a 5% chance of engaging in toxic behavior on day 366; this means $h(366) = .05$. However, this chance could increase or decrease as a function of the honesty of a particular worker, as well as her job position. The function $f(t|X)$ is then greater or less than 1 depending on the person (e.g., her level of honesty) and the situation (e.g., her particular job position) at a particular point in time. This flexible setup allows us to model a myriad of real-world settings and is the approach we will use in our estimation, as outlined below. However, first we consider some settings of the person and the situation that we can both measure in our study and expect to matter in terms of outcome.

### 2.1 Factors of the Person and the Situation

In principle, there are some identifiable factors that are likely to predict toxic behavior, especially when studying actual outcomes. Here we discuss those that we can measure in our data. The empirical proxies for these factors are discussed in section 3.1.

Job fit could be important in predicting toxic behavior. Kristof-Brown et al. (2005) find that job fit increases a worker’s strain, reduces his commitment to the organization, and increases the likelihood of him leaving. All of these factors should lead to an increased likelihood of toxic behavior.

Outside of the business ethics literature, there are also some consistent findings that overconfidence contributes to adverse behavior and outcomes. Petit and Bollaert (2011) document a set of important management and finance papers that have estab-
lished this link. Thus, toxic behavior should be increasing in a worker’s confidence, holding all else constant.

An apparently straightforward factor for measuring the propensity of misconduct is whether or not a worker agrees that the rules should always be followed. On the hand, those who follow the rules are likely to follow ethical rules as well. However, in some situations, a rule might need to be broken, even to do the "right thing." Meanwhile, another antecedent of misconduct that seems self-evident is honesty: those less honest are more likely, all things equal, to engage in misconduct. Thus, the response of an employee to an employer asking if she should always follow the rules is ambiguous. Nonetheless, it could still prove an important predictor of misconduct.

Possibly the most important factor of the work environment increasing the likelihood of misconduct involves the other workers surrounding a given worker. Pierce and Snyder (2008) find strong evidence that there are ethical-worker peer effects, akin to productivity peer effects. Thus, increased exposure to toxicity should itself lead to more toxicity.

Practically speaking, certain job positions are likely to lead to different levels of toxic behavior For example, some positions involve more regular contact with other workers. And some positions are easier to monitor than others. Other possible factors abound. Hence, the type of position a worker has, ceterus paribus, should prove an important factor.

We now turn to our estimation strategy.

2.2 Estimation Strategy

For our empirical analysis, as with our theoretical discussion, we utilize a proportional hazards model (see Cameron and Trivedi (2005)). This allows us to avoid assumptions about the shape of the base hazard rate $h(t)$ over time. We then assume that this base hazard is modified by

$$ f(t|X) = e^{\beta_p X_{p,t} + \beta_s X_{s,t}}, $$
where $x_{p,t}$ is a vector of personal traits at time $t$ and $x_{s,t}$ is a vector of situation characteristics at time $t$. The role of $e$ is simply to ensure that the composite hazard rate $h(t) f(t|X)$ is never negative.\textsuperscript{6} In other words, the baseline hazard rate $h(t)$ can be any arbitrary shape over time, but it is (potentially) modified by the person and situation at time $t$ by $f(t|X)$.

This setup then gives us the following partial log-likelihood function to maximize:

$$
\log L = \sum_{j=1}^{D} \left[ \sum_{i \in D_j} x_i \beta - d_j \log \left\{ \sum_{k \in R_j} e^{x_k \beta} \right\} \right],
$$

where $i$ indexes subjects, $x_i$ is a vector of covariates representing the person and the situation, $j$ indexes failure times in chronological order, $D_j$ is the set of $d_j$ failures at time $j$, and $R_j$ is the set of all subjects that could potentially fail at time $j$.

In our empirical setting, we have both many types of workers across workgroups and quasi-random matching to workgroups. In particular, company executives explain that the typical worker placement is a function of periodic work flow and other forces, which are not predictable. We will show in our robustness section that first placements are approximately random. In addition, we will show that our main effects persist when adding workgroup fixed effects over the first placement. However, since these robustness tests yield similar results to analysis using the full dataset with all placements, we will begin our analysis with the whole, and then turn to the parts for our robustness tests.

Since we both have a very large sample and find that our results are consistent when focusing on quasi-random placement, we also abstract away from separating the notion of engaging in toxic behavior and being terminated for toxic behavior. For expositional variety, we will use phrases such as "a worker engages in toxic behavior," "he is a toxic worker," and "she is terminated for toxic behavior" interchangeably. However, strictly speaking, these terms all mean ultimately being terminated for toxic behavior.

\textsuperscript{6}Precisely, we need the image of $f(t|X)$ to be in the set $\mathbb{R}^+ \cup \{0\}$, since it must be that $h(t) f(t|X) \in [0, +\infty)$.
behavior.

3 Empirical Analysis

We first describe our data in more detail.

3.1 Data

The data were obtained from a company that builds and deploys job-testing software to large employers. In fact, many of these companies are business-process outsourcers (BPOs) that themselves provide a variety of business services (e.g., customer care, outbound sales, etc.) to their clients. The employees included in the dataset are all engaged in frontline service positions and paid on an hourly basis. From these organizations, we were able to obtain and combine three separate datasets on the basis of employee IDs:

1) Job-testing data: The vendor supplying the data has developed a proprietary job test that assesses applicant fit for the position for which applying. We were able to obtain the employee scores on this job test (i.e., Green, Yellow, Red) as well as the responses to select questions that appeared on the test.

2) Attrition data: All of the companies with which the vendor engages provide an attrition feed that indicates (among other things) the employee’s hire date, termination date (as applicable), reason for termination, their location, job title, and the supervisors to whom they reported while employed by the firm.

3) Performance data: For a subset of employees included in our analysis, we were able to obtain daily performance data that represent productivity by measuring the average amount of time an employee required to handle a transaction and customer satisfaction scores indicating how well she served the customer.

Common employee IDs across all three of these datasets allowed us to merge them together in order to look at relationships between assessment responses and an employee’s likelihood of engaging in toxic behavior. In total, the dataset covers 11 firms, 184 sub-firms (end clients of BPOs), 3,314 workgroups, each reporting to a
particular supervisor, and 58,542 workers. Table 1 provides a summary of our main variables of interest.

From the assessment data, we were able to obtain several different measures of worker quality and predicted performance.

Each employment assessment is designed by an industrial-organizational psychologist and attempts to measure an employee’s knowledge, skills, and abilities. A proprietary scoring methodology utilized employee responses to in order to calculate a single job-fit factor score indicating their fit for the particular position (i.e., Green, Yellow, Red). Bad Job Fit is a dummy variable that takes on a 1 if the job fit score is Red, which means the assessment predicts that a worker is not a good fit for the particular job to which he is being assigned.

Also included in the assessment were questions intended to gauge an applicant’s technical ability. Applicants were asked early in the assessment to self-assess their computer proficiency and they were then tested on several key computer skills. We compared their self-assessment to their actual computer proficiency in order to develop a measure of applicant self-confidence. The variable Skills Confidence Level is constructed by extracting the residual from a regression of actual skills (i.e., measured skills) on promised skills (i.e., given by the worker). That is, this variable is a measure of how much the actual skills are above or below the promised skills.\(^7\)

We acknowledge that this variable could also be a measure of honesty. However, once hired, if a worker has lower-than-promised skills, there is a real chance that she will be terminated. Further, 11% of workers actually underpromise their skills level, which would make this an unlikely measure of dishonesty in such an incentivized setting. Finally, we still find similar results if we simply drop the 34% of workers who overpromise performance. Thus, it seems this variable is more a measure of confidence in one’s own abilities than that of honesty.

Several questions on the assessment asked applicants about their propensity to follow rules. We were able to obtain those individual questions and the applicant

\(^7\)We also calculate Skills Confidence as simply the difference between stated and actual skills without using regression analysis, and the results are similar. In absolute terms, we find that roughly 11% of workers promise lower skill than they deliver, 55% deliver as they promise, and 34% overpromise.
responses in order to understand whether there was a relationship between the response option an applicant endorsed and her likelihood to engage in toxic behavior. In particular, applicants were asked to choose one option from each of the two sets of statements:

1. I believe that rules are made to be followed. OR

2. Sometimes it’s necessary to break the rules to accomplish something

and

1. I like to see new places and experience new things. OR

2. I complete activities according to the rules.

For each of the rule-following variables constructed, a 1 means that the worker chose the statement that rules should be followed (i.e., the first statement in the first set and the second statement for second set). Thus, receiving a 1 on these dummy variables means that a subject is stating that he feels rules should be followed.

Density of Toxic Workers is a ratio that measures the degree of a worker’s exposure to other toxic workers. That is, it is the ratio of other workers on a worker’s team who are ultimately terminated for being toxic, as described below, divided by the current number of workers on the worker’s team. Thus, this measure changes over time.

For a subset of the dataset, we also have performance data. We have a measure of worker output speed. In particular, we have the length of time needed to complete one unit of output. The variable Performance Quantity Time FE is an individual worker fixed effect calculated while regressing the time-per-unit of a worker on a cubic function of time-on-the-job experience and controls for job position and the sub-firm where the worker is employed, while achieving a given performance result. That is, we generally have multiple observations of a worker’s performance over time; we refer to each observation of performance measurement as a performance result. In addition, we have a measure of worker output quality. This variable Performance Quality is obtained analogously to the variable Performance Quantity Time FE.
Finally, our dependent variable is an indicator variable based on whether the worker is terminated for toxic behavior. Toxic behavior is defined as involuntary termination due to an egregious violation of company policy. Examples include sexual harassment, workplace violence, falsifying documents, fraud, and general workplace misconduct. The mean of this variable is approximately 1% across all observations. However, in terms of per worker, the mean is 4.5% of all observations. In other words, roughly 1 in 20 workers is ultimately terminated as a toxic worker.

3.2 Hazard Functions

In this section, we report graphically overall empirical hazard rates as a function of time. In the next section, we will conduct a full analysis with controls. However, we begin by simply comparing hazard rates with our first two measures. To provide sufficient observations we report the hazards for the first 365 days, as over 90% of a worker’s tenure is under one year.

The following chart shows the difference in hazard rates of workers who are predicted to be a bad fit for the job (i.e., badscore=1) and those who are not:

As can be seen, across a worker’s employment, a worker predicted to be a bad fit
for the position is more likely to be terminated for toxic behavior, as shown by the top line.

Next, we consider the difference between above-average (i.e., $\text{conf\_level}=1$) and below-average (i.e., $\text{conf\_level}=0$) Skills Confidence Level. Those who appear overconfident by overreporting their skill level before they start the job are more likely to be terminated for toxic behavior across all time.

Although these charts are suggestive, they need to be interpreted with care; they do not include potentially important controls. Further, we need to consider different factors simultaneously to determine if these are different predictors of toxic workers or these different measures are just a measure of the same underlying force. For this analysis, we turn to our proportional-hazards regression model.
3.3 Regression Analysis

3.3.1 Baseline

Table 2 reports the results of our baseline regression model. For these regressions, we have a large enough sample to stratify by each sub-firm. This means that each sub-firm is allowed to have a unique baseline hazard function $h(t)$. That is, it is as if we estimate each regression sub-firm by sub-firm. As can be seen, greater reported Skills Confidence results in a greater chance of being terminated for being a toxic worker. In particular, a one standard deviation in Skills Confidence results in over a 16% increase in the hazard. That is, conditional on a worker not yet having been terminated as a toxic worker, a one standard deviation increase in Skills Confidence means that there is over a 16% greater relative chance of termination due to toxic behavior. If a worker reports that she believes rules are made to be followed (as opposed to stating that it is sometimes necessary to break the rules to accomplish something), she is about 25% relatively more likely to be terminated for actually breaking the rules. Finally, a worker that has a one standard deviation increase in exposure to toxic workers is himself 47% relatively more likely to be terminated for engaging in toxic behavior.

If we categorize the first four columns as measures of the person and the last two columns (including particular job type) as measures of the situation, we can state what fraction of a toxic worker’s origin is attributable to the person versus the situation. In particular, using McFadden’s pseudo $R^2$, we calculate that approximately 67% of the explanatory power of the model beyond a model with only an intercept comes from the person, and the balance (i.e., 33%) from the situation. That is, at least in our setting, there is important explanatory power in simply knowing the

---

8For all of our hazard models, we test the proportionality assumption (i.e., that the composite hazard rate is of the form $h(t)f(t)$) on the basis of the Schoenfeld residuals after fitting a given model (see Grambsch & Therneau (1994)). In all cases, our model is consistent.

9Recall that to convert estimates into a hazard ratio, simply raise $e$ to the coefficient value. For example, a coefficient value of .55 results in $e^{.55} \approx 1.73$. This means a one unit change in the regressor amounts to a 73% increase in the hazard ratio. Alternatively, a one standard deviation increase, when such standard deviation is .225, results in a roughly 16.5% increase in the hazard ratio.
person, though the situation certainly matters too.

Ideally, we would like to conduct our analysis after randomly allocating all workers to their work groups and then observing their experiences and performance over time. Doing so would average out possible confounds that are difficult to control for.

For example, perhaps a particular workgroup is better (or worse) at detecting and eliminating toxic workers. However, based on discussions with company executives, conditional on a given sub-firm, a worker’s first placement tends to be essentially random. In particular, exactly where an employee is initially placed depends on a variety of factors outside the control of the worker and the workgroup to which she is placed. For example, the work flow of a particular operation, demand and supply shocks, and exactly when a worker turns up looking for a job are all factors determining to which group a given worker will be assigned. Further, a workgroup supervisor does not generally choose her group’s worker, as a worker is hired and then assigned to a given workgroup. Also helpful is that the supervisor of the workgroup does not observe the new worker’s predicted job fit and other individual characteristic covariates that we use in our analysis. Nonetheless, a worker’s second placement may not be essentially random. Thus, for a robustness test, we now redo our above analysis, but only for a worker’s first placement.

### 3.3.2 First Placement Only

Table 3 reports analysis based only on an employee’s first placement. As can be seen, the results are broadly very similar to the case when all worker placements are included. Upon closer inspection, we see that the magnitude of the estimates on predicted job fit are slightly larger when only considering the first placement. In contrast, the magnitude of the density of other toxic workers an employee faces is much smaller. Here, a one standard deviation increase in toxic-worker exposure now results in an 11% increased hazard (versus a 47% increase when considering a worker’s exposure to toxic workers over an entire work history). This reduced magnitude is nonetheless still significant at the 1% level. One possible explanation for a smaller effect from toxic-worker exposure when only considering the first placement is that
the exposure to toxic workers has a cumulative effect: that is, the same exposure over a greater period of time has a greater adverse effect on a worker.

In principle, we can test statistically whether a placement is different from random. One common method includes comparing covariates across treatments, which is often a total of two treatments. However, in this setting, a "treatment" would be each workgroup, which amounts to 3,134 treatments, making a comparison cumbersome. Further, one can only consider relationships pair-by-pair. However, another common method that also allows the covariates to be interdependent is using a logit or probit model to predict treatment. Of course, this method only works when there are two different treatments; again, we have 3,134 treatments. However, we can analyze a multinomial equivalent where each outcome is considered an unordered outcome of being placed in a given potential workgroup. For this estimation, we need sufficient observations to estimate how each covariate might contribute to the likelihood of being placed in a particular workgroup. In addition, we should derive estimates conditional on a given subfirm, as our hazard model estimates are conditional on different subfirm-based hazards. With such an approach, we can estimate how covariates predict 938 workgroup placements.

The following table reports the results of these regressions.

<table>
<thead>
<tr>
<th></th>
<th>Bad Job Fit</th>
<th>Skills Confidence</th>
<th>Rules: Sometimes Break</th>
<th>Rules: Prefer Adventure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Workgroups Significant at 5%</td>
<td>98</td>
<td>69</td>
<td>45</td>
<td>44</td>
</tr>
<tr>
<td>Fraction of Estimated Workgroups (938)</td>
<td>10.45%</td>
<td>7.36%</td>
<td>4.80%</td>
<td>4.69%</td>
</tr>
<tr>
<td>Fraction of All Workgroups (3,134)</td>
<td>3.13%</td>
<td>2.20%</td>
<td>1.44%</td>
<td>1.40%</td>
</tr>
</tbody>
</table>

We find that in 98 of the 938 cases Bad Job fit predicts in which workgroup a worker is placed in, which represents just over 10% of workgroups. Skills Confidence is significant over 7% of the time, whereas both Rules covariates are significant less than 5% of the time. If all placements were generated at random, we would expect each covariate to be significant at the 5% level, 5% of the time, on average. The average of all four covariates is 6.82%. When we consider the full dataset we are using to estimate effects, covariates are only significant 2% of the time, on average. The rea-
son we cannot estimate covariate effects on the entire dataset is that generally there are too few observations for a particular workgroup, which means we do no expect such workgroups to be creating statistical aberrations on their own. However, there could potentially be a systematic relationship between the covariates and whether or not a worker is placed in a workgroup where effects can be estimated (i.e., one of the 938 workgroups) versus one where they cannot. To explore this possibility, we run a logit model where the treatment is whether or not the worker is placed in one of the (938) estimated workgroups. Here, Bad Job Fit is negative and significant at the 1% level. This means that those with a predicted Bad Job fit are less likely to be placed in an estimatable workgroup. Skills Confidence and Rules: Sometimes Break carry p-values of .923 and .356, respectively. Finally, Rules: Prefer Adventure is marginally significant with a p-value of .052. Thus, our two consistent covariates that matter in terms of the person in predicting toxic behavior—Rules: Sometimes and Break Skills Confidence—seem to be quasi-randomly spread across workgroups, regardless of whether we can estimate them with a multinomial model.

An additional robustness test is to explicitly control for an employee’s workgroup during her first placement. In particular, we run a linear panel model with workgroup fixed effects for a worker’s first placement. Here, we collapse the exposure to toxic workers as an average exposure over the placement, whereas before this was the current-period exposure. Results are reported in Table 4. The findings with this linear model are very similar in terms of significance compared with our hazard models. The exception is that the coefficient on a worker being a bad fit is now highly significant for all specifications. In terms of magnitude, a worker that is dubbed a bad fit has an additional 1% chance of becoming a terminated toxic worker, which is an increase of 25% from the baseline toxic worker rate of roughly 4%. A one standard deviation in Skills Confidence results in a roughly 13% chance of becoming a terminated toxic worker. Those who state that rules should never be broken are 21% more likely to be terminated for toxic behavior. Finally, a one standard deviation

\footnote{Note that we do not control for position type in these specifications. A particular workgroup typically consists of the same set of position types, and thus the variance matrix naturally becomes unusable when we do attempt to control for position type simultaneously with workgroup.}
increase in exposure to toxic workers induces a roughly 20% increased chance of a worker himself becoming a toxic worker. In short, these effects are qualitatively the same as with the previous models.

3.3.3 Toxic Worker Performance

For a subset of the data, we have performance data on the workers. For this group, we have a measure of work speed and quality. In particular, as discussed in section 3.1, we have a measure of each employee’s time to produce one unit of quantity and a measure of their quality of work. We then use this data to calculate a worker-specific fixed effect of each of these measures, which we refer to as Performance Quantity Time FE and Performance Quality FE, respectively. Table 5 reports the results of introducing these additional measures to our original analysis reported in table 2. The other variables of interest previously studied are qualitatively the same as in table 2, although the levels of significance are diminished for this considerably smaller sample size.

As far as performance, we can see that those who are terminated for engaging in toxic behavior are actually higher performers in terms of producing quantities at greater speed. That is, those who are slower (i.e., large values of Performance Quantity Time FE) are also less likely to be toxic. In terms of magnitude, a one standard deviation in time per unit of production results in a 74% reduction in the hazard of becoming a toxic worker. However, those workers with poorer quality performance are more likely to be toxic. Here, a one standard deviation increase in the quality of production results in a 12% decrease in the hazard.

Finding Superstars vs. Losing Toxic Workers With performance data we can compare the strategy of finding another "superstar" worker versus simply eliminating a toxic one. As discussed in the introduction, many firms and the extant literature are focused on finding and keeping the next star performer, whereas it seems that much less attention is devoted to limiting the chances of obtaining a new employee or pushing a current one into becoming toxic. Given a firm with limited resources,
which strategy is more fruitful? Although we certainly cannot answer this question for all possible settings, we can assess this trade-off for our setting.

To generate a straightforward comparison, we quantify the value of a star performer by identifying the cost savings from her increased output level. That is, without such a star performer, a firm would have to hire additional workers to achieve the same output when utilizing an average worker. In the below table, the column "Superstar" reports the cost saving based on the top 1%, 5%, 10%, and 25% performers. In particular, we calculate the percent in increased performance for each of these performance levels and multiply it by the average salary of an average worker.

For comparison, we then report in the "Toxic Worker" column the turnover cost of a toxic worker. Turnover cost includes both the average cost of replacing the worker, as well as increased voluntary turnover in worker teams with such a worker, based on company figures. This amounts to $12,800, $4,000 of which is based on the approximate average cost of actually replacing the toxic worker. This Toxic Worker cost does not include other potential costs, such as litigation, regulatory penalty, or reputation cost.

<table>
<thead>
<tr>
<th></th>
<th>Cost Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Superstar</td>
</tr>
<tr>
<td>top 25%</td>
<td>$1,951</td>
</tr>
<tr>
<td>top 10%</td>
<td>$3,251</td>
</tr>
<tr>
<td>top 5%</td>
<td>$3,875</td>
</tr>
<tr>
<td>top 1%</td>
<td>$5,303</td>
</tr>
</tbody>
</table>

In comparing the two costs, even if we could replace an average worker with one who performs in the top 1%, we would still be better off replacing a toxic worker with an average worker by more than two-to-one.

4 Discussion

Based on our analysis, we have a variety of takeaways for managers. From our study, it seems clear that toxic workers originate both as a function of preexisting characteristics and the environment in which they work. In particular, we found
consistent evidence that those who seem overconfident in their abilities and those who claim rules should be followed, even before being hired, are more likely to become toxic workers and break the rules. One strategy for managers is to screen potential workers for these traits to reduce the chance of hiring toxic workers. Of course, there are more dimensions to a good (and bad) hire beyond whether or not candidates have a higher propensity to become toxic. Worker productivity is also important. Interestingly, we found that toxic workers are more productive, at least in terms of the quantity of output. This could also explain how toxic workers are able to persist as much as they do. For example, an investment bank with a rogue trader who is making the firm millions in profits might be tempted to look the other way when the trader is found to be overstepping the rules. However, we also found that the quality of production for these workers is lower. This means that eventually, the value of the higher productivity will be diminished, perhaps drastically, as the consequences to the firm of lower-quality work are manifested.

Although we find certain preexisting traits that predict toxic workers, this does not mean that those traits were always present in the worker. Though it is beyond the scope of this paper, it would be interesting to learn to what extent work-life experiences develop the preexisting traits that we show lead to toxic workers. It would be very valuable to discover what firms can currently do to limit the chances of converting a "normal" worker to a future toxic worker.

In this vein, we did find that a worker’s environment, such as her particular position and set of coworkers, influenced her propensity to become a toxic worker. And to a lesser extent we found that job fit was also important. Hence, this suggests that managing toxic workers is not simply a matter of screening them out of the firm. It seems that an important aspect is also to make sure a worker is put in the most appropriate position based on her skills and predilections (Kristof-Brown et al. (2005) and Roberts and Saloner (2013)). That is, the allocation of positions seems important not just for productivity reasons but also for reducing the likelihood of spawning toxic workers. Thus, organizational design also becomes an important part of strategy in managing the corporate social responsibility of a firm: sustainability requires a sustainable workforce.
5 Conclusion

In the end, a good or bad hiring decision is multidimensional (Lazear & Oyer (2007) and Hermalin (2013)). We have identified several preexisting and job-environment factors that lead to a worker engaging in objective toxic behavior. Knowledge of these factors can be used to better manage for toxic workers. We have also discovered some important effects of toxic workers. However, there are surely additional traits that could be used to identify toxic workers. Similarly, it would be helpful to know what other environmental factors nudge an otherwise normal worker to become a toxic worker and possibly also create the preexisting conditions from the perspective of a future firm that lead to toxic behavior. Future research can shed light on these questions. This latter focus seems particularly important, because to the extent that we can reduce a worker’s propensity to become toxic, we are helping not only the firm, but the worker himself, those around him, and the potential firms where that employee may work in the future. We have taken a first step in this direction and hope that we witness future progress in this area.
References


Table 1: Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad Job Fit</td>
<td>249383</td>
<td>0.09</td>
<td>0.28</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Skills Confidence Level</td>
<td>249383</td>
<td>0.00</td>
<td>0.23</td>
<td>-0.23</td>
<td>0.92</td>
</tr>
<tr>
<td>Rules: Sometimes Break Them</td>
<td>249383</td>
<td>0.14</td>
<td>0.34</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Rules: Prefer Adventure</td>
<td>249383</td>
<td>0.44</td>
<td>0.50</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Density of Toxic Workers</td>
<td>249383</td>
<td>0.04</td>
<td>0.04</td>
<td>0.00</td>
<td>0.80</td>
</tr>
<tr>
<td>Performance Quantity Time FE</td>
<td>62618</td>
<td>-32.77</td>
<td>213.11</td>
<td>-462.94</td>
<td>1488.31</td>
</tr>
<tr>
<td>Performance Quality FE</td>
<td>20089</td>
<td>-0.05</td>
<td>0.13</td>
<td>-0.91</td>
<td>0.23</td>
</tr>
<tr>
<td>Terminated for Toxic Behavior</td>
<td>249383</td>
<td>0.01</td>
<td>0.10</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 2: Terminations as a Function of Worker Type and Environment
(All Placements)

Outcome: Terminated Toxic Worker

<table>
<thead>
<tr>
<th>Worker and Environment</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad Job Fit</td>
<td>0.2019***</td>
<td>0.1140</td>
<td>0.1017</td>
<td>0.0983</td>
<td>0.0972</td>
<td>0.1008</td>
</tr>
<tr>
<td></td>
<td>(2.91)</td>
<td>(1.61)</td>
<td>(1.43)</td>
<td>(1.39)</td>
<td>(1.37)</td>
<td>(1.42)</td>
</tr>
<tr>
<td>Skills Confidence Level</td>
<td>0.5684***</td>
<td>0.5695***</td>
<td>0.5652***</td>
<td>0.5496***</td>
<td>0.5498***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.11)</td>
<td>(7.13)</td>
<td>(7.05)</td>
<td>(6.85)</td>
<td>(6.83)</td>
<td></td>
</tr>
<tr>
<td>Rules: Sometimes Break Them</td>
<td>-0.2358***</td>
<td>-0.2246***</td>
<td>-0.2215***</td>
<td>-0.2238***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-3.72)</td>
<td>(-3.49)</td>
<td>(-3.44)</td>
<td>(-3.48)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rules: Prefer Adventure</td>
<td>-0.0349</td>
<td>-0.0263</td>
<td>-0.0251</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.86)</td>
<td>(-0.65)</td>
<td>(-0.62)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Density of Toxic Workers</td>
<td>2.5914***</td>
<td>2.5448***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(11.90)</td>
<td>(11.62)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position Controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-18477.7692</td>
<td>-18454.7262</td>
<td>-18447.4486</td>
<td>-18447.0841</td>
<td>-18398.1653</td>
<td>-18389.9353</td>
</tr>
<tr>
<td>N</td>
<td>247609</td>
<td>247609</td>
<td>247609</td>
<td>247609</td>
<td>247609</td>
<td>247609</td>
</tr>
</tbody>
</table>

Cox proportional hazard model used for estimation
Non parametric hazard functions estimated at the sub-firm level
Z scores reported in parentheses based on standard errors clustered at the worker level
* p<0.10,  ** p<0.05, *** p<.01
Table 3: Terminations as a Function of Worker Type and Environment  
(First Placements Only)

Outcome: Terminated Toxic Worker

<table>
<thead>
<tr>
<th>Worker and Environment</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad Job Fit</td>
<td>0.2327***</td>
<td>0.1351*</td>
<td>0.1235</td>
<td>0.1256</td>
<td>0.1246</td>
<td>0.1246</td>
</tr>
<tr>
<td></td>
<td>(3.10)</td>
<td>(1.76)</td>
<td>(1.60)</td>
<td>(1.63)</td>
<td>(1.62)</td>
<td>(1.61)</td>
</tr>
<tr>
<td>Skills Confidence Level</td>
<td>0.6121***</td>
<td>0.6154***</td>
<td>0.6183***</td>
<td>0.6163***</td>
<td>0.6178***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.31)</td>
<td>(6.34)</td>
<td>(6.34)</td>
<td>(6.32)</td>
<td>(6.32)</td>
<td></td>
</tr>
<tr>
<td>Rules: Sometimes Break Them</td>
<td>-0.2036***</td>
<td>-0.2099***</td>
<td>-0.2086***</td>
<td>-0.2062***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-2.76)</td>
<td>(-2.81)</td>
<td>(-2.79)</td>
<td>(-2.76)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rules: Prefer Adventure</td>
<td>0.0204</td>
<td>0.0220</td>
<td>0.0228</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.41)</td>
<td>(0.44)</td>
<td>(0.46)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Density of Toxic Workers</td>
<td></td>
<td></td>
<td></td>
<td>1.5446***</td>
<td>1.5006***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(3.98)</td>
<td>(3.82)</td>
<td></td>
</tr>
<tr>
<td>Position Controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Log Liklihood</td>
<td>-11884.4856</td>
<td>-11865.9400</td>
<td>-11861.9682</td>
<td>-11861.8839</td>
<td>-11858.1369</td>
<td>-11851.0799</td>
</tr>
<tr>
<td>N</td>
<td>191460</td>
<td>191460</td>
<td>191460</td>
<td>191460</td>
<td>191460</td>
<td>191460</td>
</tr>
</tbody>
</table>

Cox proportional hazard model used for estimation  
Non parametric hazard functions estimated at the sub-firm level  
Z scores reported in parentheses based on standard errors clustered at the worker level  
* p<0.10, ** p<0.05, *** p<.01
# Table 4: Linear Model of Terminations with Workgroup Fixed Effects
(First Placements Only)

**Outcome: Terminated Toxic Worker**

<table>
<thead>
<tr>
<th>Worker and Environment</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad Job Fit</td>
<td>0.0145***</td>
<td>0.0103**</td>
<td>0.0100**</td>
<td>0.0103**</td>
<td>0.0095**</td>
</tr>
<tr>
<td></td>
<td>(3.67)</td>
<td>(2.55)</td>
<td>(2.47)</td>
<td>(2.56)</td>
<td>(2.36)</td>
</tr>
<tr>
<td>Skills Confidence Level</td>
<td>0.0245***</td>
<td>0.0246***</td>
<td>0.0249***</td>
<td>0.0243***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.18)</td>
<td>(5.19)</td>
<td>(5.24)</td>
<td>(5.13)</td>
<td></td>
</tr>
<tr>
<td>Rules: Sometimes Break Them</td>
<td>-0.0071***</td>
<td>-0.0080***</td>
<td>-0.0078***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-2.66)</td>
<td>(-2.94)</td>
<td>(-2.88)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rules: Prefer Adventure</td>
<td></td>
<td></td>
<td></td>
<td>0.0029</td>
<td>0.0028</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.40)</td>
<td>(1.36)</td>
</tr>
<tr>
<td>Avg Density of Toxic Workers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1991***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(4.10)</td>
</tr>
<tr>
<td>R Squared</td>
<td>0.044</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
<td>0.046</td>
</tr>
<tr>
<td>Adjusted R Squared</td>
<td>0.023</td>
<td>0.023</td>
<td>0.023</td>
<td>0.023</td>
<td>0.024</td>
</tr>
<tr>
<td>N</td>
<td>45083</td>
<td>45083</td>
<td>45083</td>
<td>45083</td>
<td>45083</td>
</tr>
</tbody>
</table>

* t statistics reported in parentheses based on standard errors clustered at the workgroup level
  * p<0.10,  ** p<0.05,  *** p<.01
Table 5: Terminations with Worker Performance  
(All Placements)

Outcome: Terminated Toxic Worker

<table>
<thead>
<tr>
<th>Worker and Environment</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance Quantity Time FE</td>
<td>-0.0035*** (-6.79)</td>
<td>-0.0038*** (-5.37)</td>
<td></td>
</tr>
<tr>
<td>Performance Quality FE</td>
<td>-2.2315*** (-5.00)</td>
<td>-2.4931*** (-5.41)</td>
<td></td>
</tr>
<tr>
<td>Bad Job Fit</td>
<td>-0.0763 (-0.50)</td>
<td>-0.0903 (-0.30)</td>
<td>-0.1474 (-0.47)</td>
</tr>
<tr>
<td>Skills Confidence Level</td>
<td>0.4254*** (3.10)</td>
<td>0.4560** (2.51)</td>
<td>0.4625** (2.52)</td>
</tr>
<tr>
<td>Rules: Sometimes Break Them</td>
<td>-0.0249 (-0.23)</td>
<td>-0.0263 (-0.16)</td>
<td>0.0144 (0.09)</td>
</tr>
<tr>
<td>Rules: Prefer Adventure</td>
<td>-0.0851 (-1.22)</td>
<td>-0.2611*** (-2.64)</td>
<td>-0.2435** (-2.44)</td>
</tr>
<tr>
<td>Density of Toxic Workers</td>
<td>1.5261*** (5.19)</td>
<td>1.7699*** (5.87)</td>
<td>1.5936*** (5.17)</td>
</tr>
<tr>
<td>Position Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>yes</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-5858.5339</td>
<td>-3231.2511</td>
<td>-3163.8239</td>
</tr>
<tr>
<td>N</td>
<td>62419</td>
<td>19983</td>
<td>19751</td>
</tr>
</tbody>
</table>

Cox proportional hazard model used for estimation
Non parmetric hazard functions estimated at the sub-firm level
Z scores reported in parentheses are based on standard errors clustered at the worker level
* p<0.10, ** p<0.05, *** p<.01