Manager Feedback Style and Worker Productivity

Calvin Jahnke, Gábor Nyéki, and Carolyn Tsao

Extended Abstract

Click here for the latest version.

Abstract. This paper studies how a manager's tone when giving feedback to workers affects individual productivity and output quality. We construct a novel panel dataset that links software engineers and managers to their email communications and code contributions on the largest open source software project, the Linux kernel. We identify tones used in the emails (e.g., toxic, polite, encouraging) using natural language processing and machine learning techniques. We find a strong negative relationship between manager toxicity and engineer productivity. Using an instrumental variables design to address endogeneity in a manager's choice of tone, we find that receiving toxic feedback from a manager reduces the likelihood that an engineer completes a programming task, increases the amount of time to task completion, and decreases the likelihood that an engineer completes more tasks in the next 30 days.

Date: January 30, 2025.

1. Introduction

Managers matter for firm and worker productivity. But what specific skills or traits make for a productivity-enhancing manager? There is increasing evidence that bundled manager qualities such as leadership and organizational skills can increase productivity at the firm or team level (e.g., Syverson, 2011; Bandiera et al., 2020; Gosnell et al., 2020) and reduce turnover (e.g., Lazear et al., 2015; Hoffman and Tadelis, 2021; Alan et al., 2023), and that such skills are associated with better labor market outcomes for the managers themselves (e.g., Bertrand and Schoar, 2003; Kuhn and Weinberger, 2005). However, less is known about how specific managerial traits affect worker productivity. Furthermore, existing work on individual worker output is scarce and is primarily limited to routine occupations where output is more homogeneous and thus easier to measure.

This paper investigates the effects of one particular managerial characteristic: a manager's *feedback style*, i.e., whether they use aggressive (e.g., toxic) or non-aggressive (e.g., encouraging) language when giving feedback. Giving feedback is a core responsibility of managers in all industries, and prior work finds evidence that workers who feel that they are treated better are more productive (e.g., Krueger and Mas, 2004). However, there is virtually no empirical evidence on whether manager communication style affects worker productivity. One-on-one communication is typically not recorded or accessible, as it is exchanged either in-person or over private channels. There is also little evidence on how managers affect worker productivity in high-skill occupations such as engineering since individual output is heterogeneous and difficult to quantify.

We overcome these challenges and provide new empirical evidence on the extent to which communication style matters for worker productivity and turnover in a high-skill worker setting. We construct a novel dataset in which we observe email communications between workers and managers as well as individual output produced. We document, for the first time, the relationship between manager tone and worker outcomes, including the speed at which they complete tasks and their involvement on future tasks. We use multiple empirical strategies to credibly identify the effects of tone on worker outcomes, including an instrumental variables strategy that addresses concerns of correlations between manager tone and latent worker ability and of assignment bias caused by managers sorting to higher ability workers.

We study communication between engineers who work on the Linux kernel, the largest open source software project to date. Most engineers who work on the project do so as employees of software and hardware firms. Thus, just like software engineers in private firms (e.g., Emanuel et al., 2023), "contributors" to the Linux kernel have their code scrutinized in a code review process. Code review is primarily performed by engineers who act as "maintainers." It is their job to check, provide feedback on, and approve contributions ("patches") to the code base. Importantly, an unusual and novel feature of our setting is that all communication between maintainer–contributor pairs is publicly available. Because no single employer dominates the development of the Linux kernel, code review is done online in full transparency. Emails are the key mode of communication on the project, and all email communications include engineers' full names and firm affiliations. Moreover, feedback style is anecdotally salient in the project, as tone became a topic of lengthy controversy during the 2010s, the period that we study.

Leveraging the open source nature of the project, we construct a unique panel on contributors, maintainers, their code, and their email exchanges. We link these exchanges to LinkedIn data by name and firm affiliation to obtain information on contributor demographics, including gender, race, education, work experience, seniority at the firm, and job transitions.

We characterize each email using a discrete range of key communication tones: toxic, neutral, appreciative, and encouraging. We measure communication style and output quality using state-of-the-art measures developed in the computer science literature. To measure email tone, we treat the text in the emails as raw data and apply a host of natural language processing (NLP) techniques. Our simplest specification is an n-gram based method that indicates whether words and phrases associated with different tones appear in the emails. We also use a transformer-based model developed by Sarker et al. (2023) specifically for detecting toxicity in code reviews, a topic that has received much attention in the computer science literature.[1]

Finally, we evaluate whether tone has causal effects on contributor output. We use multiple empirical strategies to identify the causal effects of tone on two key outcomes: "patch completion," i.e., securing maintainer approval for one's submitted code and the submission of new patches in the future. Each strategy addresses a different set of threats to identification, including measurement error in tone, and potential bias in the assignment maintainers to contributors in code review. We first use an instrumental variables (IV) approach to address the possible correlation between the quality of submitted code and the tone of the maintainer. Leveraging the large number of emails per person, we instrument for a manager's tone in a particular thread with their tone in prior email

[1] Toxicity in particular has been a focus of research around the tone of code reviews in research from, for instance, Google. Their results point to the prevalence of second person pronouns in code review messages which they flag as toxic (Qiu et al., 2022). Identifying these phrases in our communications offers a point of comparison for our own tone identification as well as an additional dimension of communication: direction.

threads. However, this method is susceptible to assignment bias, and there is anecdotal evidence that some maintainers purposefully employ toxic language against contributors who submit code that they deem sub-par. We therefore use the same IV approach on small mailing lists only, which are responsible for more specialized areas of the kernel and are more likely to be associated with a single maintainer. In ongoing work, we are also implementing mover designs in which we apply the IV approach to interactions involving a first-time contributor, a newly moved contributor, or a newly moved maintainer.

Our first main finding is that contributors and maintainers alike use a wide range of tones during code review, even after conditioning on the content being discussed. The majority of emails use neutral language only (73%). Relatively few emails use outright toxic language (2%). A larger share express appreciation (6%), using phrases like "thanks" or "I really appreciate," or encouragement (6.6%), using phrases like "this is good" or "I like . . . ".

Our second main finding is that certain tones predict other tones, both in the immediate conversation as well as in future email threads. For instance, toxicity begets toxicity: those who receive a toxic message are 6% more likely to send a toxic message in reply, and 4% more likely to send a toxic message in the future on a different email thread. To a lesser degree, politeness also predicts politeness: those who receive a polite message are 8% more likely to send a polite message in reply, but are no more likely to send a polite message in the future. Gratitude and appreciation are typically only delivered in one direction (e.g., from contributor to maintainer, or vice versa), but receiving messages that convey gratitude or appreciation increases the likelihood that one uses a grateful or appreciative tone in future messages by 1.8%.

Our third main finding is that there are systematic patterns in how manager feedback tone relates to engineer productivity. Our IV estimates show a strong negative relationship between maintainer toxicity and the likelihood that a contributor completes their patch (Table 1). On the other end of the spectrum, contributors who receive an encouraging message from a maintainer have a slower than average initial response time, but are more likely to complete patches in the next 30 days. Our identification strategy, which addresses measurement error in tone as well as assignment bias of maintainers to contributors, shows larger effects than the corresponding OLS estimates. This highlights the importance of using empirical strategies that address attenuation bias in OLS estimates.

2. Related Literature

This paper contributes to a vast literature on the effects of management practices on productivity (Bertrand and Schoar, 2003; Bloom and Van Reenen, 2007; Syverson, 2011; Bandiera et al., 2020; Adhvaryu et al., 2023a). An increasing number of studies link management practices to individual-level output (Ichniowski et al., 1997; Ichniowski and Shaw, 2003; Bloom et al., 2013; Gosnell et al., 2020; Hoffman and Tadelis, 2021; Alan et al., 2023) or team-level output (Chen, 2021; Weidmann and Deming, 2021; Fenizia, 2022; Adhvaryu et al., 2023b). Much of the related literature focuses on the effects of management practices on the output of rote work. For example, Krueger and Mas (2004) find that plant workers provide "more effort and due diligence" if they feel they are treated better. However, more recent studies have begun to tackle either the effects of managerial "soft skills" or the output of highly-skilled labor. For example, Hoffman and Tadelis (2021) find that managers with better people management skills reduce worker turnover and appear to be rewarded by the firm for these skills, and Gosnell et al. (2020) find that "pertinent" management practices can also impact the productivity of more highly-skilled labor (namely, airline captains).

Our paper tackles both margins—by studying the output of highly-educated software engineers *and* implementing new measures of communication skills—and in doing so, contributes further evidence to the nascent literature that soft skills matter for worker output.

References

Adhvaryu, A., N. Kala, and A. Nyshadham (2023a): "Returns to On-the-Job Soft Skills Training," *Journal of Political Economy*, 131, 2165–2208.

Adhvaryu, A., A. Nyshadham, and J. Tamayo (2023b): "Managerial Quality and Productivity Dynamics," *The Review of Economic Studies*, 90, 1569–1607.

Alan, S., G. Corekcioglu, and M. Sutter (2023): "Improving workplace climate in large corporations: A clustered randomized intervention," *The Quarterly Journal of Economics*, 138, 151–203.

Bandiera, O., A. Prat, S. Hansen, and R. Sadun (2020): "CEO behavior and firm performance," *Journal of Political Economy*, 128, 1325–1369.

Bertrand, M. and A. Schoar (2003): "Managing with style: The effect of managers on firm policies," *The Quarterly journal of economics*, 118, 1169–1208.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013): "Does management matter? Evidence from India," *The Quarterly journal of economics*, 128, 1–51.

Bloom, N. and J. Van Reenen (2007): "Measuring and explaining management practices across firms and countries," *The quarterly journal of Economics*, 122, 1351–1408.

Chen, Y. (2021): "Team-Specific Human Capital and Team Performance: Evidence from Doctors," *American Economic Review*, 111, 3923–62.

Emanuel, N., E. Harrington, and A. Pallais (2023): "The power of proximity to coworkers: training for tomorrow or productivity today?" Tech. rep., National Bureau of Economic Research.

Fenizia, A. (2022): "Managers and Productivity in the Public Sector," *Econometrica*, 90, 1063–1084.

Gosnell, G. K., J. A. List, and R. D. Metcalfe (2020): "The impact of management practices on employee productivity: A field experiment with airline captains," *Journal of Political Economy*, 128, 1195–1233.

Hoffman, M. and S. Tadelis (2021): "People Management Skills, Employee Attrition, and Manager Rewards: An Empirical Analysis," *Journal of Political Economy*, 129, 243–285.

Ichniowski, C. and K. Shaw (2003): "Beyond incentive pay: Insiders' estimates of the value of complementary human resource management practices," *Journal of Economic Perspectives*, 17, 155–180.

Ichniowski, C., K. Shaw, and G. Prennushi (1997): "The effects of human resource management practices on productivity: A study of steel finishing lines," *American Economic Review*, 87, 291–313.

Krueger, A. B. and A. Mas (2004): "Strikes, Scabs, and Tread Separations: Labor Strife and the Production of Defective Bridgestone/Firestone Tires," *Journal of Political Economy*, 112, 253–289.

Kuhn, P. and C. Weinberger (2005): "Leadership skills and wages," *Journal of Labor Economics*, 23, 395–436.

Lazear, E. P., K. L. Shaw, and C. T. Stanton (2015): "The value of bosses," *Journal of Labor Economics*, 33, 823–861.

Qiu, H. S., B. Vasilescu, C. Kästner, C. Egelman, C. Jaspan, and E. Murphy-Hill (2022): "Detecting interpersonal conflict in issues and code review: cross pollinating open- and closed-source approaches," 41–55.

Sarker, J., A. K. Turzo, M. Dong, and A. Bosu (2023): "Automated Identification of Toxic Code Reviews Using ToxiCR," ArXiv:2202.13056 [cs].

Syverson, C. (2011): "What determines productivity?" *Journal of Economic literature*, 49, 326–365.

Weidmann, B. and D. J. Deming (2021): "Team Players: How Social Skills Improve Team Performance," *Econometrica*, 89, 2637–2657.

Table 1. Baseline Results of Maintainer Toxicity on Contributor Outcomes

	First Stage (1)	OLS (2)	IV (3)	Reduced Form (4)
	A. Time to first response (hours)			
Maintainer's avg. toxicity on other patches	1.55***			-22.04
	(0.06)			(29.96)
Maintainer toxicity on current patch		-3.95**	-14.21	
		(1.74)	(19.31)	
N	109,021	109,021	109,021	109,021
First stage F-stat	741.03			
Mean of Dependent Variable	0.03	29.01		
	B. Likelihood of patch completion			
Maintainer's avg. toxicity on other patches	1.26***			-1.81***
	(0.03)			(0.10)
Maintainer toxicity on current patch		-0.07***	-1.44***	
		(0.01)	(0.08)	
N	244,798	244,798	244,798	244,798
First stage F-stat	1,439.39			
Mean of Dependent Variable	0.02	0.46		
	C. Time to patch completion			
Maintainer's avg. toxicity on other patches	0.66***			59.00*
	(0.06)			(34.45)
Maintainer toxicity on current patch		3.02	89.96*	
		(2.34)	(53.07)	
N	76,960	76,960	76,960	76,960
First stage F-stat	131.02			
Mean of Dependent Variable	0.02	30.22		
	D. Likelihood of future patch completion (next 30 days)			
Maintainer's avg. toxicity on other patches	1.23***			-0.31***
	(0.03)			(0.05)
Maintainer toxicity on current patch		-0.01	-0.25***	
		(0.00)	(0.04)	
N	214,987	214,987	214,987	214,987
First stage F-stat	1,229.56			
Mean of Dependent Variable	0.02	0.51		

Note: The sample in each regression above includes messages (i) from maintainers (ii) that are in response to patches (iii) that are not their own. All regressions include controls for contributor (employee), list, and year fixed effects, along with the number of files changed, line insertions and deletions of the patch, the number of patches previously submitted by both the maintainer and the contributor, and whether the patch was submitted by another maintainer. In Panel A, the dependent variable is the number of hours before the first response to the maintainer's code review message. In Panel B, the dependent variable is a dummy variable for whether the patch is ever completed. In Panel C, the dependent variable is the number of days between the initial patch submission and patch completion. In Panel D, the dependent variable is the fraction of a contributor's patches submitted in the next 30 days (excluding resubmissions of the current patch) that become commits. * Significant at 10% level, ** Significant at 5% level, *** Significant at 1% level.