



Rational sycophants and catastrophic risks

Robert Axelrod^{a,1} and Scott E. Page^{a,b,1,2}

Contributed by Scott E. Page; received December 29, 2025; accepted April 10, 2026; reviewed by John H. Aldrich and Margaret Levi

Sycophants praise and support leaders' proposals to gain personal and professional advantage. A rational sycophant is an advisor who supports actions they expect to be harmful even when rewards and punishments for good and bad advice are equal in magnitude. Rational sycophancy arises when the outcome distribution of a proposed action has a negative expected value but a positive median (outcome asymmetry). The risk is greatest when a small yet meaningful fraction of outcomes are catastrophic, which occurs in long-tailed distributions. Given that single realizations from long-tailed distributions reveal little about the underlying distribution, even after outcomes are observed, a leader may be unable to distinguish rational sycophancy from wise counsel. As a result, rational sycophants may gain influence and increase the likelihood of catastrophic policy outcomes.

sycophancy | decision-making | catastrophic risk

Presidents, prime ministers, and autocrats as well as leaders of large organizations—corporate CEOs, university presidents, and heads of nonprofits, rely on experts to provide advice on high-stakes decisions. Ideally, these advisors act as honest brokers and speak truth to power. However, advisors may also want to maintain their access to power to influence future policy choices. These considerations create incentives for sycophancy: supporting a leader's opinions and proposals even when those ideas lack merit or, worse, have potentially damaging consequences.

Ideally, a leader relies on decision-making processes, creates an organizational culture, and implements incentive structures to prevent sycophancy. Here, we present a logic for why that may be more difficult than previously thought. We demonstrate that a rational advisor may support a policy proposal that they expect to be harmful if that advisor is rewarded and punished equally for good and bad advice. This rational sycophancy is possible only if the distribution of policy outcomes satisfies an asymmetry condition in which the median outcome has a positive value, and the mean outcome has a negative value.

In our model, we assume a single advisor who has perfect information about the distribution of outcomes. To the extent that the advisor cares about policy outcomes, their preferences are identical to those of the leader. As an example, an advisor who predicts that a proposed policy will produce a benefit of X with probability 90% and losses of twenty times X with probability 10%, would expect to receive a reward for giving good advice nine times as often as receiving the equally sized punishment. If the advisor cares little (or not at all) about the policy's effects on society, that advisor will support the policy even though the mean value is negative.

As we discuss later, rational sycophancy can also arise in a team of advisors who possess different, and imperfect information as well as heterogeneous preferences. Neither form of diversity overcomes the potential for rational sycophancy.

A rational sycophant tells the leader what the leader wants to hear based on explicit cost–benefit calculations. This differs from a traditional (psychological) sycophant who is someone prone to the use of flattery to curry favor or gain influence.

Concern that sycophancy undermines political decision-making dates back to Athenian democracy (1, 2). It continues to this day. Political leaders generally choose their cabinet members and senior advisors, whereas in the corporate and nonprofit sectors board members often either select or approve senior leaders. This makes political organizations more prone to sycophancy. Leaders such as Xi Jinping (3), Vladimir Putin (4, 5), and Donald Trump (6) have all been characterized as overly reliant on sycophants.

A leader surrounded by sycophants will not hear dissenting voices. Such a leader may make poor decisions or take no action at all. Evidence shows that by increasing leader confidence, sycophants make leaders less inclined to take necessary strategic actions (7). Research on *AI* sycophancy, the tendency of *AI* to provide answers people desire, has also been shown to promote overconfidence (8).

Significance

Political and organizational leaders rely on advisors to make decisions. Ideally, leaders select honest advisors and not sycophants who praise and support the leader to gain personal or professional advantage. We show that avoiding sycophancy may be more difficult than previously thought. Advisors who receive rewards and punishments of equal magnitude for good and bad advice can be rationally sycophantic. They can support proposed actions that they expect to be harmful. Rational sycophancy arises when policy outcomes satisfy outcome asymmetry: Their outcome distributions have negative expected values but positive medians. We then show why long-tailed outcome distributions, which arise in policy domains such as financial markets, international conflicts, ecological systems, and epidemics, may pose the greatest risk from rational sycophancy.

Author affiliations: ^aFord School of Public Policy, University of Michigan, Ann Arbor, MI 48104; and ^bRoss School of Business, Santa Fe Institute, Santa Fe, NM 87501

Author contributions: R.A. and S.E.P. wrote the paper.

Reviewers: J.H.A., Duke University; and M.L., Stanford University.

The authors declare no competing interest.

Copyright © 2026 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹R.A. and S.E.P. contributed equally to this work.

²To whom correspondence may be addressed. Email: scottepage@gmail.com.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2537944123/-/DCSupplemental>.

Published May 26, 2026.

Many historical examples of failed decisions are partly explained by sycophantic advisors not challenging leaders. Examples include the Challenger disaster, the failed US war effort in Vietnam (9), IBM's failure to enter the personal computer market, Volkswagen's decision to approve tampering with exhaust systems to meet emissions standards, and GM's decision to stick with a flawed ignition switch design (10).^{*} In each case, advisors who were aware of potentially harmful outcomes did not speak out. Our model demonstrates that support for these and other failed decisions may have been rational given the advisors' incentives.

We view rational sycophancy as potentially more dangerous, as a leader may be aware that it occurs. In our models, we show how rational sycophancy can produce catastrophic outcomes with nontrivial probabilities. These catastrophic outcomes are more likely in the modern interconnected world in which actions by state and nonstate actors can disrupt the stability of large systems—financial markets, power grids, information infrastructure, and ecosystems. States can support or allow research that could release pathogens resulting in a pandemic (11).

Rational Sycophancy

To formally define rational sycophancy, we first define traditional sycophancy as an advisor who supports a proposed policy or action with an expected loss of L . For this to happen, the sycophant must receive value for supporting the leader that exceeds their concern for social loss. The value that accrues to the sycophant could have any of three sources. The sycophant could value offering positive support, value receiving praise from the leader, or wish to avoid punishment for opposing a proposal. In sum, a traditional sycophant values giving and receiving praise (and dislikes punishment) more than they care about policy outcomes.

We define rational sycophancy as different from the traditional conception. As shown in Fig. 1, a rational sycophant earns a reward (punishment) R ($-R$) if the proposed policy is implemented and proves successful (unsuccessful). If the advisor opposes a policy, we assume they receive no payoff.[†]

Let p denote the probability that the policy will be successful. A rational sycophant can support a proposal with an expected social loss, L , provided their expected reward, $pR + (1 - p)(-R)$, minus any reduction in utility they receive from the social loss, is positive. We let θL denote the reduction in the payoff to the advisor from the social loss, so that when $\theta = 0$, the sycophant only cares about their private reward. Rational sycophancy requires an asymmetric outcome distribution in that most outcomes have positive values, but the expected value of the proposed action must be negative.

Returning to our initial example, the proposed policy has an expected value of minus $0.1X$. A rational sycophant would support the policy provided that their expected private reward, $0.8R$, exceeds their loss in utility from the policy being implemented, $\theta(0.1X)$.

^{*}Volkswagen's sophisticated "defeat device" escaped detection by regulators for six years.

[†]We do not include payoffs to the advisor for not supporting a proposed policy. Doing so would require an additional assumption about whether the leader follows the advice. This would complicate the model without adding new insights. A possible extension would be to have the payoff from rejecting the policy to have an asymmetric outcome distribution. Our core logic would still apply. Advocating the status quo would also produce positive (negative) rewards depending on the outcome and the advisor may choose the more dangerous alternative.

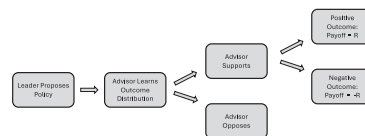


Fig. 1. A Model of a Rational Sycophant. The leader proposes a policy but does not know the distribution of outcomes. The advisor learns the outcome distribution and then either supports or opposes the policy. If the advisor supports the policy, the policy is implemented and the advisor earns a reward (punishment) of R if the policy is successful (unsuccessful).

Asymmetric Outcome Distributions and Sycophancy. Rational sycophancy requires that the distribution of outcomes satisfy two conditions. First, it must have a positive **median**; most outcomes must be successful. Second, it must have a negative **mean**; the expected value to society must be negative.[‡]

Our rational sycophant model differs from a standard principal-agent model in which an advisor (the agent) has private knowledge, information, or expertise and strategically craft messages to produce decisions more to their liking than what would occur from full information revelation (12). In principal-agent models, the agent prefers a policy some distance from the true state, while the principal wants a policy set equal to the true state. The agent distorts the final policy by sending a common signal for a range of values of states of the world resulting in a policy biased toward the agent's preferred point. If the agent makes the pooling interval too large, the principal's action will be far from the true state of the world, and then both the agent and the principal will get low payoffs. The central finding is that policy distortions exist, though they tend not to be large. We should not expect catastrophes.

That will not be true in our model. We do not assume that the advisor knows the true state of the world or that the advisor gets a more accurate signal. Instead, we assume that the advisor knows the distribution of outcomes of a proposed policy but that the leader does not. That informational advantage allows the advisor to calculate the expected value of a proposed policy and the probability of a positive outcome.

The key differences between our model and a principal-agent model merit repeating. In principal-agent models, the agents manipulate the principal's information to obtain a policy more to the agent's liking. In our model, sycophants who maximize their payoffs given a reward structure support policies with negative expected value—and potentially catastrophic outcomes. Our rational sycophants are not altering policies.

Long-Tails and Catastrophic Risk

The expected losses from rational sycophancy depend on the difference in the proportion of successful and unsuccessful outcomes. As the probability of success increases, two results follow: i) the magnitude of expected losses that a rational sycophant would support increases and ii) the magnitude of the worst possible outcome becomes unboundedly large as the probability of success approaches one.

We derive these two results by generalizing our example. We assume that the proposed policy produces a benefit of X with probability p and with probability $(1 - p)$ it produces a loss of kX , where $k > 1$. Recall that a rational sycophant supports

[‡]From the inequality above, rational sycophancy requires

$$p \cdot R + (1 - p) \cdot (-R) - \theta L > 0$$

This expression can be written as $(2p - 1) \cdot R > \theta L$, implying that p must exceed $\frac{1}{2}$.

a policy provided $(2p - 1)R + \theta(pX - (1 - p)kX) > 0$. The minimal reward to support the policy $\bar{R}(p)$ as a function of the probability of a successful outcome can be written as follows:

$$\bar{R}(p) = \frac{\theta((1 - p)kX - pX)}{2p - 1}$$

This function has a negative derivative. As p increases, the numerator decreases, and the denominator increases. Therefore, holding $\bar{R}(p)$ fixed, if we increase p , we increase the numerator, that is increase k , to obtain the same minimal reward. Larger values of k imply a lower expected value of the outcome proving the first result.

The second result follows from the same logic. The bad outcome has magnitude kX . As p approaches one, the policy becomes almost certain to be successful. Given that $(1 - p)$ approaches zero, to keep $\bar{R}(p)$ unchanged, k must approach infinity.[§]

Long-Tailed Distributions

As just shown, the most extreme sycophantic risk occurs for distributions that result in successful outcomes most of the time, yet still have negative expected values because of nonnegligible likelihoods of catastrophic outcomes. These two features are present in long-tailed outcome distributions, such as lognormal and power law distributions.

To build intuition, we begin with the familiar case of a lognormal distribution. Suppose that a policy has a known gross benefit, $B = 10$, and a lognormal distribution that characterizes losses, given by $LN(\mu, \sigma^2)$. If we set $\mu = 0$, the mean of the distribution equals $e^{\frac{\sigma^2}{2}}$ whereas the median equals one.[¶] If $\sigma = 3$, then for approximately 88% of outcomes, the policy will produce a positive net benefit, that is, losses less than ten. Expected losses, though, exceed ninety.

Lognormal distributions have moderately long tails. And, as just shown, if policy losses are drawn from a lognormal distribution, then rational sycophancy can result in risky policy choices. Far more concerning are policies in which the potential losses are drawn from a power law (Pareto) distribution which could have a much longer tail.

We place particular emphasis on power law distributions because they occur empirically in many policy domains: sizes of military conflicts, market collapses, power grid failures, social movements, and natural disasters. All have been shown to be well approximated by power law distributions (13, 14).

Power law distributions can arise for a variety of reasons (15–17). Most relevant to our analysis is that power law distributions arise in domains in which actions spread across networks through positive feedbacks such as in power outages and bank failures. A failure of one power station or bank places stress on neighbors in the network, and, when those fail as well, a cascade ensues.

These domains are particularly relevant because a political leader may not be well versed in the technical intricacies of power grids, financial markets, artificial intelligence, or epidemics. Similarly, a corporate leader may possess a deep understanding of their market or industry, but not understand the potential impact

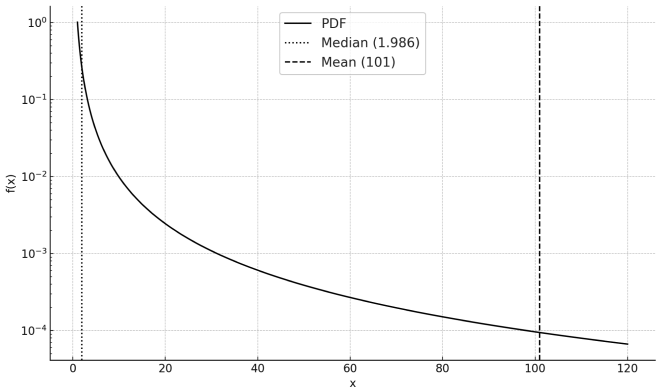


Fig. 2. A Probability Density Function (PDF) for a power law distribution $f(x) = (a - 1)x^{-a}$ with $a = 2.01$ plotted on log scale. The distribution has a mean value of 101 and a median of 1.986. If we interpret this as the cost distribution to a policy with a benefit greater than two but less than one hundred and one, then the policy has a negative expected value but a positive median value.

of disruptions such as artificial intelligence, cryptocurrencies, or a green energy transition.

A power law distribution of the form $f(x) = (a - 1)x^{-a}$ has a mean value of $(a - 1)/(a - 2)$ and a median of $2^{\frac{1}{a-1}}$. As with a lognormal distribution, the median is strictly less than the mean (SI Appendix). For power law distributions, the divergence of the mean from the median can be extreme. As a approaches two, the mean increases without bound, yet at $a = 2$, the median equals only two.

Fig. 2 shows a power law density function with $a = 2.01$. This distribution satisfies the outcome asymmetry condition in extreme form. The mean equals 101, and the median equals two. Given this cost distribution, if the gross benefit equals twenty, more than 95% of outcomes would be positive, yet the expected value would be negative eighty-one.[#]

Similar to our analysis of the lognormal distribution, we assume a proposed policy has a known gross benefit, which we take to be $B + 1$. We assume the net effect of the proposal to equal this benefit minus a cost drawn from a power law distribution with exponent $a \in (1, 2)$. We can thus write the net outcome, y , as $y = (B + 1) - x$. The expected net outcome equals $B + 1 - (a - 1)/(a - 2)$, which can be simplified as $B - 1/(a - 2)$.

a	Mean	Median	Pr(outcome < −100)
2.01	−97	2.01	1%
2.5	1	2.41	0.1%

An advisor, the rational sycophant, receives a reward (punishment) $R > 0$ ($-R$) if the proposal has a positive (negative) value. Let p denote the probability that the outcome is positive. We assume that the rational sycophant’s payoff equals their value from the social benefit ($\theta > 0$ times the mean policy outcome), plus the probability of a good outcome times their reward minus the probability of a bad outcome times their punishment. Thus, the advisor supports a proposal if and only if

[§]A symmetric argument reveals an implication of our model. A rational advisor could also be a *rational contrarian* and oppose a proposal with a positive expected value if most of the time the proposal produces a negative outcome. An advisor might argue against holding a peace summit knowing the summit has little chance of success, even if, in those cases where the summit proves successful, the social benefit is enormous.

[¶]Note that the extent of outcome asymmetry, the ratio of the mean (losses) to the median, increases exponentially in σ .

[#]Power laws have much higher probabilities of large events than normal distributions. The right tail of a normal distribution decreases like e^{-x^2} . If costs are normally distributed, with a mean of two and a SD of one, an event with costs greater than eight occurs once in 500 million trials. For a power law distribution, $f(x) = (a - 1)x^{-a}$ with $a = 3$, the distribution also has mean of two, yet one in every sixty-four outcomes have costs greater than eight, and one in every ten thousand outcomes has costs exceeding one hundred.

$$\theta \left(B - \frac{1}{a-2} \right) + (2p-1)R \geq 0$$

The above expression reveals that the payoff to the advisor decreases with the reciprocal of $(a-2)$, so that as a approaches two, the advisor's expected value decreases without bound. If the advisor places sufficient weight on the social benefit, that is if θ is relatively large, the advisor will not support such a proposal. However, if θ is small or zero, a rational advisor could support proposals in which the exponent is near two. In these cases, the advisor places almost no weight on the social value. They care almost entirely about the probability of a good outcome. Thus, they could support policies with extremely negative expected values so long as p exceeds one-half, which is true for power law distributions with exponents near two.

One might expect that a leader could learn that an advisor was a sycophant and avoid that risk. That inference holds when outcomes are drawn from normal distributions. In normal environments, a leader can identify an advisor with a low concern for the public interest and a large interest in rewards. In fact, the leader would likely be able to make such inferences after just a handful of policy outcomes.

In contrast, when costs are drawn from a power law distribution, the leader can learn little about the advisor's θ . A single realization or even a moderate number of realizations provides little information (*SI Appendix*). Thus, rational sycophants can persist because they will be difficult to identify.^{||}

To demonstrate the difficulty of this inference problem, we first present a numerical example that provides the core intuition. We set $B+1$ equal to four creating a mean outcome value of $4 - (a-1)/(a-2)$. Assume that losses are equally likely to be drawn from either of two power law distributions: a high-risk distribution ($a = 2.01$), a low risk distribution.

The high-risk distribution is ten times more likely to produce a catastrophic outcome (losses greater than 100). Yet, these two distributions' medians differ by less than one-half. For most outcomes, a leader would not be able to infer much about the distribution.

We show this formally in two ways (*SI Appendix*). First, we apply a maximum likelihood approach to show that it takes one million draws from a power law distribution to get similar accuracy that one would obtain from ten draws from a normal distribution. Second, we apply Bayesian classification methods to solve for the probability of a correct classification as a function of the difference in means for normal distributions and power law distributions. As shown in Fig. 3, for a normal distribution with a SD equal to one, when the difference in means exceeds six, a single draw can distinguish with near certainty which was the true distribution. For two power law distributions, the difference in their means can equal eighty, and one draw will nearly always provide little information about which is the true distribution. As shown in Fig. 3, classification accuracy barely exceeds 50%, which corresponds to a random guess (*SI Appendix*).

Teams of Advisors

Our analysis so far considers a single advisor. In most policy contexts, leaders rely on multiple advisors. Sycophancy can still arise in teams through social or strategic conformity. Social conformity arises when people agree with the group rather

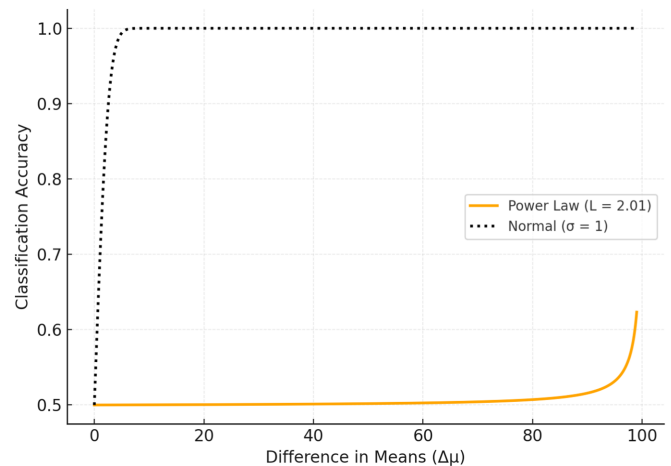


Fig. 3. Classification accuracy as a function of the difference in the mean for two normal distributions with SD equal to one and for two power law distributions in which one of the distributions has exponent $a = 2.01$.

than voice their own opinion in order to maintain status or be liked (19). The proportion of people who conform due to social pressure is often overstated (20). In Asch's (21) famous conformity experiments approximately two-thirds of responders gave the correct answer despite social pressure. Nevertheless, pressure to conform could mean that a team of advisors all align with a sycophant.

In teams, strategic conformity could amplify rational sycophancy. Strategic conformity occurs when someone acts against their own information after updating on information presented by others. If advisors make binary recommendations and do so sequentially, an information cascade can arise. A collection of advisors could, on average and correctly, believe a policy to be so risky that not even a rational sycophant would approve. However, if the first advisors to weigh in on the policy perceive the least risk, the policy could be approved (22). If advisors were to provide substantive reasons for their positions, they might prevent cascades.

Decision-making protocols can amplify sycophancy in teams by encouraging conformity. Unanimity, whether enforced by rule or social convention, creates both social and strategic incentives that can result in groupthink (23). If team members interpret going along with the crowd as loyalty, poor decisions follow (24). Contrary to intuition, if actors are strategic and care only about the correctness of the decision, unanimity may more often lead to bad decisions than majority rule (25). A strategic actor has incentives to support the policy because the only case in which their vote matters is if everyone else supports the proposal. And, conditional on being in that situation (being pivotal), an advisor can be more likely to be wrong. So, they should vote against their information.

Even if a leader selects advisors who differ in how much they value private rewards and punishments, that diversity will not prevent rational sycophancy if advisors gain or lose influence based on their successes and failures (*SI Appendix*). Any sequence of good outcomes on risky proposals would increase the influence of rational sycophants with less social concern (lower θ). The median advisor would shift to someone sycophantic resulting in the potential for riskier and riskier proposals being approved. Catastrophe becomes not just possible, but probable. The apt metaphor here is not simply a leader who takes big risks—who tosses matches into a dry forest—but a leader whose advisors

^{||}Taleb et al. (18) make an even stronger comment about point forecasts from long-tailed distributions, describing them as "unscientific simplifications" and "close to useless."

support taking on bigger and bigger risks, who support tossing matches into drier and drier forests over time.

In general, a leader who gives accurate advisors more influence will reduce diversity. On binary decisions, a group is only necessarily more accurate than its average member if its members make highly diverse classifications (26). If a leader only makes proposals they think have positive net value, this narrows the set of proposals advisors see. It follows that advisors who think similarly to the leader will be more likely to approve and, given our assumptions, increase in influence. And, if the advisors rely on a small set or similar models to make policy evaluation, they will be collectively less accurate (27). In sum, rewarding success with teams reduces diversity and encourages sycophancy resulting in even worse policy advice and greater risk of catastrophe.

Conclusion

In this paper, we show that rational sycophancy can arise when policies generate asymmetric outcome distributions, and advisors receive equal rewards and punishments for good and bad advice. Of course, only the first of these is necessary. Furthermore, so long as there exists a bound on maximal punishment for bad advice, the possibility of rational sycophancy exists even with punishments that exceed rewards.

The potential risks of sycophancy depend on the extent of the outcome distribution; how much the mean outcome is less than the median. In contexts in which outcome distributions have long (negative) tails, and in which leaders are limited in the extent to which they can reward and punish advisors, catastrophes may arise.

In our analysis, we did not distinguish between democracies and autocracies, the logic suggests that rational sycophancy could arise in both. It may be more prevalent in autocracies where leaders have less access to independent information, and where leaders can stay in power for decades. An advisor who expects many more opportunities to influence policies has greater incentive to support a policy that produces good outcomes most of the time (3). Thus, we would expect sycophancy to be more prevalent in strong autocracies given their longer duration (28). Autocrats can stay in power for decades.

Our analysis suggests multiple ways in which political systems can reduce the likelihood of rational sycophancy. First, and most important, leaders should avoid rewards and punishments based on outcomes. Leaders would be better served by evaluating advisors based on the depth and coherence of their input. Such guidance can be difficult to follow. Leaders cannot help but reward advisors who give advice that turns out well.

Leaders often take calculated risks by placing trust in their advisors. In our model, rational advisors prove untrustworthy owing to the asymmetry of outcomes. To further reduce sycophancy, leaders can put in place formal decision-making protocols that include consideration of worst case scenarios (29, 30). These protocols can also change the culture so that advisors are predisposed to be honest.

Institutional rules that require the approval of multiple decision makers can also reduce the risk of bad decisions. In democratic systems, approval by a legislature provides a check on rational sycophancy (31). This would be especially true in divided governments where “red teaming”—challenging the assumptions of the other side, is a common practice.

International bodies, such as the United Nations can raise awareness of potential risks and convene leaders and experts to raise opposition to dangerous policies as can a free press. Both can motivate people to organize against dangerous policies or to take individual actions to mitigate potential catastrophic effects, such as building air raid shelters, sheltering in place, or not using artificial intelligence.

We have placed less emphasis on the potential firms. Historically, the scale of risk from poor business decisions is smaller than in the government setting. This is not to minimize losses to human life from chemical poisonings, plant explosions and meltdowns, poor automobile designs, building collapses, adverse health outcomes from pharmaceutical innovations, or to understate the enormous potential risks associated with artificial intelligence.

Data, Materials, and Software Availability. There are no data underlying this work.

1. J. O. Lofberg, *Sycophancy in Athens* (Collegiate Press, Menasha, WI, 1917).
2. D. J. D'Amico, The law and economics of sycophancy. *Const. Polit. Econ.* **29**, 424–439 (2018).
3. J. Blanchette, Xi Jinping's faltering foreign policy. *Foreign Aff.* **March 16** (2022).
4. A. Batur, N. Khokhlov, J. Tolstrup, Playing the sycophant card: The logic and consequences of professing loyalty to the autocrat. *Am. J. Polit. Sci.* **69**, 1180–1195 (2025).
5. A. Libman, Excessive Loyalty in Putin's Authoritarian Regime: The Costs of Sycophancy. (Springer), (2025).
6. A. Fiala, *Tyranny from Plato to Trump: Fools, sycophants, and citizens* (Bloomsbury Publishing PLC, 2022).
7. S. H. Park, J. D. Westphal, I. Stern, Set up for a fall: The insidious effects of flattery and opinion conformity toward corporate leaders. *Adm. Sci. Q.* **56**, 257–302 (2011).
8. J. Kwik, Digital yes-men: How to deal with sycophantic military ai? *Glob. Policy* **16**, 467–473 (2025).
9. H. R. McMaster, *Dereliction of duty* (HarperCollins, New York, 1997).
10. T. Eifler, T. J. Howard, The importance of robust design methodology: Case study of the infamous gm ignition switch recall. *Res. Eng. Des.* **29**, 39–53 (2018).
11. R. Brooks, Understanding current threats to democracy: The limits of the civil-military relations paradigm. *Daedalus* **149**, 231–257 (2020).
12. S. Gailmard, J. W. Patty, *Learning While Governing: Expertise and Accountability in the Executive Branch* (University of Chicago Press, Chicago, 2012).
13. A. Clauset, C. R. Shalizi, M. E. J. Newman, Power-law distributions in empirical data. *SIAM Rev.* **51**, 661–703 (2009).
14. C. Cioffi-Revilla, M. I. Midlarsky, Power laws, scaling, and fractal patterns in the severity of armed conflicts. *J. Conf. Resolut.* **57**, 683–713 (2013).
15. M. E. J. Newman, Power laws, pareto distributions and zipf's law. *Contemp. Phys.* **46**, 323–351 (2005).
16. P. Bak, *How Nature Works: The Science of Self-Organized Criticality* (Copernicus, New York, 2013).
17. S. E. Page, *Diversity and Complexity* (Princeton University Press, Princeton, NJ, 2011).
18. N. N. Taleb, Y. Bar-Yam, P. Cirillo, On single-point forecasts for fat-tailed variables. *Int. J. Forecast.* **38**, 1358–1362 (2022).
19. E. E. Jones, *Ingatiation: A social psychological analysis* (Appleton-Century-Crofts, New York, 1964).
20. R. Friend, Y. Rafferty, D. Bramel, A puzzling misinterpretation of the asch conformity study. *Eur. J. Soc. Psychol.* **20**, 29–44 (1990).
21. S. E. Asch, Studies of independence and conformity: I. a minority of one against a unanimous majority. *Psychol. Monogr. Gen. Appl.* **70**, 1–70 (1956).
22. S. Bikhchandani, D. Hirshleifer, I. Welch, A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Polit. Econ.* **100**, 992–1026 (1992).
23. I. L. Janis, *Victims of groupthink* (Houghton Mifflin, Boston, 1972).
24. A. O. Hirschman, *Exit, Voice, and Loyalty* (Harvard University Press, Cambridge, MA, 1970).
25. T. J. Feddersen, W. Pesendorfer, The swing voter's curse. *Am. Econ. Rev.* **86**, 408–424 (1996).
26. L. Hong, S. E. Page, The range of collective accuracy for binary classifications under majority rule. *Econ. Theory* **79**, 275–300 (2025).
27. E. P. Economou, L. Hong, S. E. Page, Social structure, endogenous diversity, and collective accuracy. *Journal of Economic Behavior & Organization* **125**, 212–231 (2016).
28. P. F. Andersson, J. Teorell, The double-edged sword: how state capacity prolongs autocratic tenure but hastens democratization. *J. Conf. Resolut.* **69**, 980–1004 (2025).
29. G. Klein, Performing a project pre-mortem. *Harv. business review* **85**, 18–19 (2007).
30. J. Grusin, S. Lindo, *Intelligent Analysis: How to Defeat Uncertainty in High-Stakes Decisions*. (Intelligent Risk Management Publications), (2021).
31. J. Bednar, *The robust federation: Principles of design* (Cambridge University Press, 2008).