Games and Economic Behavior 131 (2022) 245–263

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Information manipulation and competition ☆

Andreas Grunewald [a], Matthias Kräkel [b,*]

[a] Frankfurt School of Finance and Management, Department of Economics, Adickesallee 32-34, 60322 Frankfurt, Germany
[b] University of Bonn, Institute for Applied Microeconomics, Adenauerallee 24-42, D-53113 Bonn, Germany

ARTICLE INFO

Article history:
Received 4 March 2020
Available online 3 December 2021

JEL classification:
H00
D02
D72
D80
L10

Keywords:
Backfiring
Competition
Information manipulation
Signal jamming

ABSTRACT

In the last decade, social media and the Internet have amplified the possibility to circulate false information among an audience, which has become a serious threat to the credibility of politicians, organizations, and other decision makers. This paper proposes a framework for investigating the incentives to strategically manipulate the audience's belief under different institutions and in various competitive environments. We show that more rigorous institutions against information manipulation can lead to higher manipulation intensities in equilibrium. Complementary, we study what kind of competitive environment is particularly susceptible to the manipulation of information.

1. Introduction

The rapid advance of social media seems to have induced an increase in the dispersal of false information in a large variety of applications (Allcott and Gentzkow, 2017; Lazer et al., 2018).[1] In political elections, fake news are spread to manipulate voters' beliefs in a favored direction (Kang, 2016; Beck and Witte, 2017; Sanger, 2017), or simply because individuals unintentionally circulate erroneous information.[2] There are manipulated product reviews on internet forums (Harmon, 2004; Mayzlin et al., 2014; Zinman and Zitzewitz, 2016), lobbyists make up fictitious arguments to steer a decision makers' policy, and workers spread false information to improve their career prospects (Duffy et al., 2002; Murphy, 1992). The systematic manipulation of information can undermine the reputation of entire groups of important actors in our society such as politicians, firms and executives. While information manipulation is prevalent and occurs in seemingly unrelated applications, it is still poorly understood which competitive environments are particularly susceptible to this issue and how to design institutions to reduce information manipulation in the media.

☆ Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1 – 390838866. Financial support by the DFG, grant KR 2077/3-1, is gratefully acknowledged.

* Corresponding author.
E-mail addresses: a.grunewald@fs.de (A. Grunewald), m.kraekel@uni-bonn.de (M. Kräkel).

[1] "Fake news spread on social media is one of the "biggest political problems facing leaders around the world", says Jim Messina, a political strategist who has advised several presidents and prime ministers." (The Economist, November 4th, 2017, p. 21.)

[2] See, e.g., Shane (2017) and Sanger (2017) on Russian fake news to influence the 2016 presidential election in the US, Kang (2016) and Hsu (2017) on fake news websites that accused a pizza restaurant of being the home base of a child abuse ring led by Hillary Clinton and her campaign chief, John D. Podesta, and Beck and Witte (2017) and Oltermann (2017) on dirty campaigning via fake news during the Austrian election 2017.

https://doi.org/10.1016/j.geb.2021.11.007

The main contribution of this paper is twofold. On the one hand, we study how institutions against information manipulation, e.g., a public agency or a law that forces media companies to delete false information, affect players' manipulation incentives. We show that more rigorous institutions might backfire and end up in boosting the overall extent of manipulation. In particular, more rigorous institutions let new information appear more trustworthy to recipients, who therefore rely more heavily on news. As a consequence, players' manipulation incentives can be aggravated. On the other hand, we investigate how the competitive environment – henceforth simply referred to as *environment* – influences incentives to manipulate information. For this purpose, we analyze a game with an audience and two competing players, and focus on two kinds of environment. First, we consider an environment in which the players' payoffs from winning or losing are constant prizes. In the second environment, payoffs from winning or losing depend on the margin of the victory, i.e., by how much the winner outperformed the loser. By these two environments, we aim to capture the two leading payoff structures in many applications. For example, constant payoffs are often attributed to presidential elections, markets with price insensitive consumers, or promotion tournaments. In contrast, environments in which the margin of the victory is important are used to model parliamentary elections (see Fishburn and Gehrlein, 1977; Grossman and Helpman, 1996) or markets with vertical product differentiation (see Shaked and Sutton, 1982). Comparing the incentives to manipulate information across environments then generates insights on how the type of the election or the structure of the market affects manipulation of information.

Our model setup postulates two assumptions on how manipulation takes effect. First, each player has an uncertain quality such that there is room for manipulating the audience's beliefs about the players' quality. Second, the audience observes a quality signal about each player, which summarizes information from newspapers, television, the Internet, and other media. While the signal is partly informative, players may manipulate its content by the costly creation of false information, e.g., by hiring a blogger who posts on the Internet. Information manipulation is a hidden action by the players, i.e., the audience cannot directly observe whether quality signals are distorted. Thus, technically, our approach belongs to the economic literature on signal jamming. After observing the – possibly manipulated – quality signals, the audience updates its beliefs about the players' qualities. In the following, we will refer to the player with the higher posterior expected quality as the player with a *posterior lead*, whereas the player with the initially higher expected quality at the beginning of the game is labeled as the player with the higher *prior lead*. We assume that in both environments the player with the posterior lead wins the competition, but only in the second environment the players' payoffs depend on the magnitude of the posterior lead. A common interpretation of the posterior lead is the better player's reputational advantage so that in either environment the competitor with the higher reputation wins, but in the second environment the players' reputations also determine the magnitude of their payoffs.

Our first set of results compares equilibrium manipulation intensities for a constant payoff structure (first environment) with those for a payoff structure that depends on the magnitude of the posterior lead (second environment). Which kind of environment implies more manipulation is an obvious question of interest not only because both kinds are frequently observed, but also because the kind of environment is often at the discretion of a superior organization – e.g., the form of an election is determined in constitutions. We show that the effect of the payoff structure on manipulation intensities crucially depends on the initial degree of player heterogeneity in terms of the magnitude of the prior lead.

If the prior lead is small, there will be a balanced competition under a constant payoff structure so that both players have strong incentives to invest in information manipulation. As a consequence, overall manipulation will be higher compared to an environment in which both players' payoffs depend on the magnitude of the posterior lead. The opposite is true if the prior lead is large. In this case, a constant payoff structure implies that both players' incentives to manipulate information are negligible since the winner of the competition is essentially predetermined by the large prior lead. If payoffs depend on the magnitude of the posterior lead, however, at least one player has incentives to manipulate information, irrespective of the initial degree of heterogeneity. Overall, the degree of initial heterogeneity, thus, determines which payoff structure leads to more information manipulation.

Our second set of results describes how the players' manipulation intensities are affected by institutions that foster transparency in the media, i.e., institutions like public agencies that erase fake reviews or hinder the publication of false statistics. We derive the following main finding: Irrespective of the payoff structure, more rigorous institutions will **increase** at least one player's manipulation intensity if and only if players are sufficiently heterogeneous. In an environment where the posterior lead is relevant, only the manipulation intensity of the initially trailing player is aggravated. Under constant payoffs, however, both players will in fact spread more false information as a response to more rigorous institutions if initial heterogeneity is large. Intuitively, if players differ strongly in their prior quality, incentives for manipulating information are low as it is almost certain who will become the player with the posterior lead. More rigorous institutions, however, let the received signals appear more trustworthy to the audience, which restores overall incentives to manipulate information.

The third set of results discusses to what extent our findings are robust to alternative ways of modeling institutions against information manipulation. To this end, we show that the key assumption in our model is that institutions cannot or do not discriminate between false information that is intentionally spread by the players and false information from other sources. If this assertion is true, more rigorous institutions let the public information appear more trustworthy to the audience, which may aggravate incentives to manipulate information. We show that this pattern arises for different kinds of institutional setup, for example if institutions provide additional information instead of eliminating false pieces of information. If, in contrast, institutions solely eliminate the manipulation by the two players, more rigorous institutions reduce players' manipulation intensities in equilibrium.

In our model, players choose hidden intensities to manipulate the quality beliefs of an audience. From a technical perspective it is therefore closely related to the literature on signal jamming, which studies a manager's borrowing decision (Stein, 1989), a worker's effort decision (Meyer and Vickers, 1997; Holmström, 1999; Höffler and Sliwka, 2003), and the use of propaganda by a dictator (or an incumbent regime), who wants to sustain his power (Little, 2012, 2017; Edmond, 2013). To analyze how manipulation intensities are affected by institutions and the competitive environment, our framework distinguishes itself from the previous contributions in two important aspects. First, we consider a setting in which two players (instead of one) choose manipulating actions. We can therefore compare for a given application (e.g., elections) how different forms of competition affect manipulation intensities. As argued in Miklós-Thal and Ullrich (2015) and Grunewald and Kräkel (2017), settings featuring competition are also qualitatively different from settings with individual decisions as the typical signal-jamming effects (e.g., higher quality uncertainty leads to more signal jamming) can be overturned.[3] Second, we study institutions against the presence of false information and their effect on manipulation intensities. While we model institutions in a reduced form, the corresponding insights can also be helpful to understand how policies against hidden and wasteful actions can backfire in settings with only one player.

Building upon a model with signal jamming allows us to construct a tractable model in which two competing players manipulate public signals at some costs. A common alternative approach to model the strategic design of information is Bayesian persuasion (see Kamenica and Gentzkow, 2011; Kamenica, 2019). This approach typically assumes full commitment of a single sender to an information structure and costless signals in the spirit of Crawford and Sobel (1982). The assumption of costless communication is almost universal in the literature but can be alleviated for single sender games for some classes of cost functions (see Gentzkow and Kamenica, 2014). For the case of costless communication, the setting has also been extended by various contributions to the case of multiple senders (see for example Gentzkow and Kamenica, 2016, 2017). Having competition and costly manipulation of information combined in a tractable fashion allows us to study how payoff structures as well as institutions affect the intensity of costly information manipulation.

A key assumption in our setup is that players manipulate information because they care about their perceived quality. In particular, the outcome of the competition does not directly depend on players' true quality but rather on the belief about their quality of a third party. We share this assumption with the literatures on reputational cheap talk (Levy, 2004; Ottaviani and Sørensen, 2006a), and signal jamming (Holmström, 1999). The assumption reflects the idea of career concerns and is commonly made, for example, in the context of political elections (Little, 2012, 2017; Grunewald et al., 2020b), professional forecasting (Ottaviani and Sørensen, 2006b; Anbarci et al., 2017), and workers' effort decisions (Holmström, 1999; Prat, 2005). Career concerns are often argued to be beneficial for a third party because they can motivate individuals to increase their effort (Holmström, 1999). In contrast, in our setting career concerns motivate players to manipulate public information with potentially detrimental consequences. We share the view that career or reputational concerns can also hinder communication with Morris (2001) and Ely and Välimäki (2003) who show that reputational concerns can even completely evade information transmission.

Our findings that more institutional scrutiny of information transmission can backfire if players can manipulate public signals is reminiscent of a number of findings showing that it is sometimes optimal to limit transparency. In a setting with one worker, Prat (2005) has shown that transparency about the worker's action might evade a principal's ability to extract information.[4] In a moral-hazard setting with multiple workers, Dubey and Wu (2001) and Dubey and Haimanko (2003) have shown that it is optimal for a principal to control the performance of workers only at a small number of instances. Such a policy will allow the principal to keep rewards small while ensuring high effort provision. The most important distinction between these two papers and ours is that they analyze a moral-hazard setting. Hence, institutional scrutiny is intended to motivate agents to exert productive effort but there is no underlying ability or quality which an audience seeks to generate information about.

Finally, our paper is related to the theoretical work that discusses incentive schemes relying on the relative performance of players. While the literature on labor market tournaments (e.g., Lazear and Rosen, 1981; Nalebuff and Stiglitz, 1983) and forecasting contests (e.g., Ottaviani and Sørensen, 2006b; Banerjee, 2021) typically consider settings in which players' payoffs from losing and winning are constant, there are also papers on the performance evaluation between career concerned experts for whom the posterior lead is important (e.g., Andina-Díaz and García-Martínez, 2021). By considering how the competitive environment affects manipulation incentives of players and the effectiveness of institutions, our results complement the above papers that take the nature of the competitive environment as given.

2. The model

We consider a competition between two risk-neutral players, A and B. The two players face an audience and simultaneously choose their intensities of manipulating information in order to improve their expected qualities from the audience's point of view. We use q_i as a measure of player i's quality ($i = A, B$). This quality is incompletely known to both players

[3] While both of these papers also study signal jamming in a framework with competition, their focus differs from the focus of this paper. Miklós-Thal and Ullrich (2015) analyze the effect of ability uncertainty on effort provision in promotion tournaments and Grunewald and Kräkel (2017) the impact of product features on advertising. Neither of them study the role of institutions or how signal jamming differs across environments.

[4] Relatedly, Anbarci et al. (2017) show in a reputational cheap talk setting that it may sometimes be optimal to obscure the information of an expert about the state of the world to optimally learn about the talent of this expert.

and the audience, who share prior beliefs about q_i.[5] In particular, we assume that q_i is distributed according to a normal distribution with mean $\bar{q}_{i0} > 0$ and variance σ_{i0}^2.[6] This lack of precise information reflects that players' quality is often a matter of taste and therefore depends on the players' personal characteristics as well as the audience's preferences. As long as no individual is perfectly informed about both of these objects, the overall quality match will thus be uncertain.

After the players chose their manipulation intensities, the audience observes a quality signal s_i about each player i. Signal s_i summarizes, for example, information from newspapers, television shows, and online forums. On the one hand, the signal reveals information on the player's true quality, q_i. On the other hand, the signal can be distorted for two reasons. *First*, players can endogenously manipulate it by creating false information. Player i can, for example, hire a blogger to create positive news about himself on the Internet.[7] These actions lead to an advantageous manipulation of the audience's beliefs about the player's quality. Player i's chosen manipulation intensity is denoted by $f_i \geq 0$,[8] which is privately observed by player i. *Second*, the quality signal can be distorted for exogenous reasons, as there might exist false information that does not stem from the players' activities. For example, articles in the public press and on websites may unintentionally be based on erroneous information, or individuals that are not associated with one of the players intentionally spread false information and rumors to enhance their own degree of popularity. All these exogenous distortions and manipulations are captured by the random variable φ_i, which is assumed to be normally distributed with $\varphi_i \sim N\left(0, \sigma_\varphi^2\right)$.

One major goal of the paper is to study the impact of institutions that foster market transparency by reducing fake news and other informational distortions in the media (e.g., a public agency or a law that forces media companies to delete false information). Crucially, we think of such institutions to reduce all kinds of false information in the media to the same extent. In particular, institutions will not be able to discriminate between intentionally and unintentionally posted fake news or between false information from the players themselves and third parties who are not directly involved in the game. To model such institutions in reduced form, we introduce a parameter $\beta > 0$ that determines how strongly $f_i + \varphi_i$ affects the audience's quality signal[9]:

$$s_i = q_i + \beta \cdot (f_i + \varphi_i) \quad (i = A, B). \tag{1}$$

Lower values of β correspond to more rigorous institutions. In the limit case of perfectly rigorous institutions (i.e., $\beta \approx 0$), the audience receives precise information about a player's true quality q_i. However, if institutions are lenient, it might even be the case that false information affects the quality signal more strongly than the underlying quality (i.e., $\beta > 1$).

Manipulating information with intensity f_i leads to costs $c(f_i)$ for player i with $c(0) = c'(0) = 0$, $c'(f_i) \to \infty$ for $f_i \to \infty$, and $c'(f_i), c''(f_i) > 0$ for $f_i > 0$, i.e., the more intensely i invests in information manipulation the higher will be his costs. The cost function is supposed to reflect various kinds of cost. In particular, spreading false information may cause immediate costs for hiring a blogger or an organization but also delayed costs as information manipulation may increase the probability of legal prosecution and potential compensation payments. Moreover, we assume that c'' is bounded from below with $c'' \geq \underline{c} > 0$. All random variables are assumed to be statistically independent.

After having observed s_i, the audience updates its prior beliefs about the distribution of qualities. From DeGroot (1970) we know that Bayesian updating conditional on s_i leads to a posterior distribution $q_i \sim N\left(\bar{q}_{i1}, \sigma_{i1}^2\right)$ with $\bar{q}_{i1} = \bar{q}_{i0} + \Sigma_i \cdot (s_i - \beta \hat{f}_i - \bar{q}_{i0})$ and $\sigma_{i1}^2 = \Sigma_i \cdot \beta^2 \sigma_\varphi^2$, where

$$\Sigma_i := \frac{\sigma_{i0}^2}{\beta^2 \sigma_\varphi^2 + \sigma_{i0}^2} \tag{2}$$

describes the prior variance of quality q_i relative to the variance of the quality signal s_i.[10] The variable $\hat{f}_i$ denotes the audience's belief about player i's manipulation intensity.[11]

[5] The informational assumptions of our model are typical of the signal-jamming literature; see, e.g., Holmström (1999). While the assumption that players do not know their qualities simplifies the analysis considerably, it should not be crucial for our results. Edmond (2013) shows in a model that combines signal jamming and private information about types that a unique equilibrium still exists. The major complication of private information is that the updating procedure of the audience needs to account for the possibility that manipulation intensities depend on the players' qualities. However, the intuition of our results does not immediately depend on the exact form of the updating procedure and should therefore be robust to slight adaptions of it.

[6] If the quality of one of the players is known, this player will not invest in information manipulation – unless he can negatively influence the signal on the other player's quality – but for the player with the unknown quality all our results remain to hold.

[7] As each player's payoff is determined by the relative comparison of both players' qualities from the audience's point of view in the following, it is not necessary to differentiate between false information that positively influences own perceived quality and false information that negatively influences the opponent's perceived quality.

[8] While it appears natural that manipulation intensities are positive, we can also imagine situations in which players would like to spread bad information about themselves. None of the results derived below hinge on the assumption that $f_i \geq 0$.

[9] Following the signal-jamming literature – e.g., Holmström (1999), Meyer and Vickers (1997), and Stein (1989) – we use a linear signal structure in our setting.

[10] Following DeGroot (1970), the posterior distribution can be derived by applying Bayes' rule. In particular, it can be shown that the posterior density of a given quality q_i is proportional to its prior density times the density of the observed signal realization conditional on q_i being the true quality. Using basic properties of the density of normal distributions then yields the result.

[11] Note that we assume the audience to hold a point belief. As we will study pure-strategy equilibria below, this assumption will necessarily hold in equilibrium.

We aim at considering manipulation intensities across different competitive environments. Therefore, at this point, we impose only mild assumptions on the competition and specify its exact structure in Sections 4.1 and 4.2 below. However, to derive some general results, we only consider environments such that the following two assumptions hold:

(i) The payoffs u^i of player $i \in \{A, B\}$ can be written as a piecewise function of the posterior lead, i.e.,

$$u^i = \begin{cases} u_L(\bar{q}_{j1} - \bar{q}_{i1}) & \text{if } \bar{q}_{i1} < \bar{q}_{j1} \\ u_H(\bar{q}_{i1} - \bar{q}_{j1}) & \text{if } \bar{q}_{i1} \geq \bar{q}_{j1} \end{cases}$$

(ii) $u_H(\cdot) : \mathbb{R}_0^+ \to \mathbb{R}$ and $u_L(\cdot) : \mathbb{R}^+ \to \mathbb{R}$ are continuously differentiable and the player with posterior lead obtains a higher payoff, i.e., $u_H(x) > u_L(x)\ \forall x > 0$.

Thus, each player's payoff depends on his relative expected quality after the audience updated its quality beliefs against the background of the received signals.

As a solution concept, we apply pure-strategy perfect Bayesian equilibrium. Hence, an equilibrium of the game consists of a pure-strategy profile incorporating the strategies of both players and a belief system such that the following two statements hold. First, both players play mutually best responses, anticipating the audience's Bayesian updating. Second, on the equilibrium path the audience and the players derive their quality perceptions from players' equilibrium manipulation choices.

3. Manipulating information

In this section, we derive the manipulation intensities chosen by the players in equilibrium. Recall that player $i \in \{A, B\}$ earns payoff $u_H(\bar{q}_{i1} - \bar{q}_{j1})$ if he becomes the player with the posterior lead and $u_L(\bar{q}_{j1} - \bar{q}_{i1})$ otherwise. He thus maximizes

$$E[u_H(\bar{q}_{i1} - \bar{q}_{j1}) | \bar{q}_{i1} > \bar{q}_{j1}] \cdot P\left(\bar{q}_{i1} > \bar{q}_{j1}\right) + E[u_L(\bar{q}_{j1} - \bar{q}_{i1}) | \bar{q}_{i1} < \bar{q}_{j1}] \cdot P\left(\bar{q}_{i1} < \bar{q}_{j1}\right) - c(f_i), \tag{3}$$

where $P\left(\bar{q}_{i1} > \bar{q}_{j1}\right)$ denotes the probability that i will be the player with the posterior lead and E the expectation operator with respect to q_A, q_B, φ_A and φ_B.

The key variable determining payoffs and, hence, incentives to manipulate information is the posterior lead from the perspective of the audience. This lead is composed of stochastic and deterministic terms. Recall that $\hat{f}_i$ denotes the audience's belief about the manipulation intensity by player i, and define by $\tilde{f}_j$ the belief that player i holds about the manipulation intensity of player j. For an easier comprehension of the problem, we separate out the stochastic elements of player i's expectation over the posterior lead $\bar{q}_{i1} - \bar{q}_{j1} = \delta_i - \Psi_i - \beta\Sigma_j(\tilde{f}_j - \hat{f}_j) + \beta\Sigma_i(f_i - \hat{f}_i)$, with δ_i being stochastic and Ψ_i embracing the exogenous deterministic elements:

$$\delta_i := \Sigma_i \cdot (q_i + \beta\varphi_i) - \Sigma_j \cdot \left(q_j + \beta\varphi_j\right) \qquad \Psi_i := \left(1 - \Sigma_j\right)\bar{q}_{j0} - (1 - \Sigma_i)\,\bar{q}_{i0}. \tag{4}$$

Since any convolution of two normal densities again yields a normal density (e.g., Ross, 2010, pp. 35, 67–68), the composed random variable δ_i is normally distributed: $\delta_i \sim N\left(\mu_{\delta_i}, \sigma_{\delta_i}^2\right)$ with

$$\mu_{\delta_i} := \Sigma_i\bar{q}_{i0} - \Sigma_j\bar{q}_{j0} \tag{5}$$

$$\text{and} \quad \sigma_{\delta_i}^2 := \Sigma_i^2\sigma_{i0}^2 + \Sigma_j^2\sigma_{j0}^2 + (\Sigma_i^2 + \Sigma_j^2)\beta^2\sigma_\varphi^2 = \Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2. \tag{6}$$

Let g_i denote the density of δ_i, and G_i the corresponding cumulative distribution function. Consequently, player i's objective function (3) can be rewritten as

$$\int\limits_{\Psi_i + \beta\Sigma_j(\tilde{f}_j - \hat{f}_j) - \beta\Sigma_i(f_i - \hat{f}_i)}^{\infty} u_H\left(\delta_i - \Psi_i - \beta\Sigma_j(\tilde{f}_j - \hat{f}_j) + \beta\Sigma_i(f_i - \hat{f}_i)\right) g_i(\delta_i)\, d\delta_i$$
$$+ \int\limits_{-\infty}^{\Psi_i + \beta\Sigma_j(\tilde{f}_j - \hat{f}_j) - \beta\Sigma_i(f_i - \hat{f}_i)} u_L\left(\Psi_i + \beta\Sigma_j(\tilde{f}_j - \hat{f}_j) - \beta\Sigma_i(f_i - \hat{f}_i) - \delta_i\right) g_i(\delta_i)\, d\delta_i - c(f_i).$$

Solving for pure strategy Perfect Bayesian Equilibria of the game leads to the following result:

Proposition 1. *In any interior equilibrium in pure strategies, player i's equilibrium manipulation intensity, f_i^*, is described by*

$$\beta\Sigma_i\left[(u_H(0) - u_L(0))\, g_i(\Psi_i) + \int\limits_{\Psi_i}^{\infty} u_H'(\delta_i - \Psi_i)\, g_i(\delta_i)\, d\delta_i - \int\limits_{-\infty}^{\Psi_i} u_L'(\Psi_i - \delta_i) g_i(\delta_i)\, d\delta_i\right] = c'\left(f_i^*\right). \tag{7}$$

Proposition 1 provides two insights with respect to the extent of manipulation in different environments and different institutional configurations. First, the payoff structure determines the marginal payoffs $u_H'(\cdot)$ and $u_L'(\cdot)$, which will shape players' manipulation incentives (see (7)). In particular, equilibrium intensities will crucially depend on whether players face a constant payoff structure so that $u_H'(\cdot) = u_L'(\cdot) = 0$, or whether the posterior lead is payoff relevant with $u_H'(\cdot), u_L'(\cdot) \neq 0$.

Second, Proposition 1 shows that the overall effect of more rigorous institutions on manipulation intensities is composed of two sub-effects. To distinguish between them, let the term in square brackets in (7), which reflects the kind of environment, be denoted by $\mathcal{E}_i$. Then, the effect of institutions on manipulation intensities can be partitioned in the following way:

$$\underbrace{\frac{\partial\ \beta\Sigma_i}{\partial\beta}\cdot\mathcal{E}_i}_{\text{information effect}} \quad + \quad \underbrace{\beta\Sigma_i\cdot\frac{\partial\ \mathcal{E}_i}{\partial\beta}}_{\text{environment effect}} \quad . \tag{8}$$

The *information effect* describes how a change in the institutional setup affects players' manipulation intensities holding the influence of the environment fixed. It arises because institutions shape how the audience reacts to newly arriving information. Therefore, it is reflected by the impact of β on $\beta\Sigma_i$, where Σ_i denotes the weight by which the audience updates its prior belief when observing the quality signal. However, the change in the audience's perception of the signals also induces a second effect, which differs across environments—the *environment effect*. It describes how the impact of the change in the audience's perception of the signals' trustworthiness differs across environments. In other words, the change in the updating weights Σ_i and Σ_j change the incentives to manipulate information differently across environments, which is captured by the environment effect.

As (7) shows, $\mathcal{E}_i$ is positive in each interior equilibrium of the game, irrespective of the environment. Hence, the direction of the information effect is independent of the environment and given by

$$\frac{\partial}{\partial\beta}\beta\Sigma_i \quad = \quad \Sigma_i + \beta\frac{\partial}{\partial\beta}\Sigma_i \quad = \quad \Sigma_i - \frac{2\beta^2\sigma_\varphi^2\Sigma_i}{\beta^2\sigma_\varphi^2 + \sigma_{i0}^2}. \tag{9}$$

On the one hand, Σ_i is always positive. The more rigorous institutions are – i.e., the smaller β – the weaker will be the manipulation of the signal, which renders false information less effective and reduces the players' incentives to manipulate information. On the other hand, the second part is negative reflecting that more rigorous institutions induce the audience to rely more heavily on the news it receives – i.e., the weight for Bayesian updating, Σ_i, becomes larger – and information manipulation becomes more appealing. The second part will dominate the first part whenever $\beta^2\sigma_\varphi^2 > \sigma_{i0}^2$. Hence, the information effect of more rigorous institutions will increase players' inclination to create false information if most of the variance of the signal s_i stems from exogenous distortions and not from quality uncertainty. In this case, the signal that the audience observes is not informative and will be almost disregarded. More rigorous institutions then restore the signal's credibility, which may lead to a higher manipulation intensity.

Importantly, the effect of the increase in Σ_i will differ across environments, i.e., the sign and the magnitude of the impact of more trustworthy signals might be affected by the structure of the players' payoffs. While the information effect of more rigorous institutions will be a recurrent theme across different environments, the sign of the environment effect is ambiguous. In the following, we will show under which conditions the environment effect even dominates the information effect such that backfiring can occur although the information effect is positive.

4. The influence of the environment on information manipulation

We start by analyzing the equilibrium manipulation intensities for two classes of competitive environments. We will then leverage the derived equilibrium intensities to address the question what kind of environment is more susceptible to the manipulation of information.

4.1. Competition with a constant payoff structure

As a first environment, suppose that u_H and u_L are exogenously given constants with $u_H = \bar{u}_H > \bar{u}_L = u_L$. This competitive environment captures situations in which the player with a posterior lead receives a high fixed benefit that does not depend on the magnitude of this lead, e.g., an election in which the election winner receives a fixed benefit. For the case of constant payoffs, Proposition 1 yields the following result:

Corollary 1. *Suppose that $u_H = \bar{u}_H$ and $u_L = \bar{u}_L$ with $\bar{u}_H > \bar{u}_L$. There will exist a unique pure-strategy equilibrium $\left(f_A^*, f_B^*\right)$ if for both players*

$$-\beta^2\Sigma_i^2\left(\bar{u}_H - \bar{u}_L\right)g_i'\left(\Sigma_i\bar{q}_{i0} - \Sigma_j\bar{q}_{j0} + \sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}\right) < \underline{c}. \tag{10}$$

In this equilibrium, players choose manipulation intensities $f_i^* > 0$ being described by

$$\beta \Sigma_i(\bar{u}_H - \bar{u}_L) g_i(\Psi_i) = c'\left(f_i^*\right) \tag{11}$$

with $i, j = A, B$ *and* $i \neq j$.

While inequality (10) states a sufficient condition for the existence of interior pure-strategy equilibria, equation (11) describes the manipulation behavior of both players in equilibrium. The equation shows that the environment effect is determined by the influence of the updating weights Σ_i and Σ_j on player i's marginal winning probability in equilibrium, $g_i(\Psi_i)$ (see (4)–(6)). These updating weights influence the magnitude of the factor Ψ_i as well as the shape of the density $g_i(\cdot)$ because both mean and variance are functions of Σ_i and Σ_j.

In addition, equation (11) shows that the manipulation incentives of both players will be small if the prior lead $|\bar{q}_{j0} - \bar{q}_{i0}|$ is sufficiently large. In particular, $\Psi_i = \mu_{\delta_i} + \bar{q}_{j0} - \bar{q}_{i0}$, and $g_i(\cdot)$ tends to zero when the absolute value of its argument grows large. Intuitively, a large prior lead describes a situation with strong initial heterogeneity among the players, so that the corresponding uneven competition only generates poor incentives.

4.2. Competition with an affine payoff structure

The previous section has analyzed manipulation intensities if players face a constant payoff structure. However, in many environments a player's payoff from winning (losing) a competition is not only determined by having a higher (lower) expected quality. Often, payoffs additionally depend on the magnitude of a player's relative expected quality compared to his opponent, e.g., the power of a politician after having won an election typically depends on the size of his victory. This section characterizes the players' manipulation intensities in equilibrium in such cases.

To model cases where the posterior lead is payoff relevant, we assume that u_H and u_L are affine functions such that player i receives the high payoff

$$u_H(\bar{q}_{i1} - \bar{q}_{j1}) := \bar{\eta} + \eta_H \cdot (\bar{q}_{i1} - \bar{q}_{j1}) \tag{12}$$

if becoming the strong player with posterior lead $\bar{q}_{i1} - \bar{q}_{j1} > 0$, whereas he gets the low payoff

$$u_L(\bar{q}_{j1} - \bar{q}_{i1}) := \bar{\eta} + \eta_L \cdot (\bar{q}_{j1} - \bar{q}_{i1}), \tag{13}$$

if becoming the weak player with $\bar{q}_{i1} < \bar{q}_{j1}$, where $\eta_H \geq |\eta_L|$ and $\eta_H > 0$. While we assume this linear structure of payoffs with identical constants for simplicity, the applications in Section 7 clarify that exactly this structure arises endogenously in many well-known theoretical settings.[12] The relation $\eta_H \geq |\eta_L|$ implies that the strong player is, at least weakly, better off than the weak player. To capture a large variety of different theoretical settings, we allow η_L to be positive as well as negative. If $\eta_L > 0$, it still holds that each player aims to be the one with the posterior lead. However, the player that is in the weaker position prefers to be of low expected quality. Such situations typically arise if two firms invest in (deceptive) advertising and strictly benefit from vertical product differentiation, i.e., from products that maximally differ in quality, to alleviate price competition (see, e.g., Shaked and Sutton, 1982). If, in contrast, $\eta_L < 0$, each player prefers to have a high expected quality irrespective of his quality ranking. Such a situation is typical of parliamentary elections, where also the inferior party benefits from a large number of received votes, which determine the party's number of seats in parliament. Combining (12) and (13) with Proposition 1 yields the following corollary:

Corollary 2. *Suppose* u_H *and* u_L *are affine and described by (12) and (13). There will exist a unique pure-strategy equilibrium* $\left(f_A^*, f_B^*\right)$ *if for both players*

$$\beta^2 \Sigma_i^2 \left(\eta_H + \eta_L\right) g_i\left(\mu_{\delta_i}\right) < \underline{c}. \tag{14}$$

The equilibrium is characterized by a corner solution $f_i^* = 0$ *if and only if* $(\eta_H + \eta_L) G_i(\Psi_i) \geq \eta_H$. *Otherwise, players choose manipulation intensities* $f_i^* > 0$ *being described by*

$$\beta \Sigma_i \left[\eta_H - (\eta_H + \eta_L) G_i(\Psi_i)\right] = c'\left(f_i^*\right) \tag{15}$$

with $i, j = A, B$ *and* $i \neq j$.

Corollary 2 shows how an affine payoff structure affects players' incentives to create false information.[13] In equilibrium, player i's probability of becoming the player with a posterior lead is given by $P(\bar{q}_{i1} - \bar{q}_{j1} > 0) = P(\delta_i > \Psi_i) = 1 - G_i(\Psi_i)$. Thus, equation (15) shows how this probability determines i's manipulation intensity in equilibrium. If this probability is

[12] While the derivations turn out to be technically involved when substantially generalizing the payoff structure, our results qualitatively transfer to slightly different parametrizations of payoff functions (say, quadratic payoffs).

[13] In order to ensure that the manipulation intensities are interior in both environments, conditions (10) and (14) and $(\eta_H + \eta_L) G_i(\Psi_i) < \eta_H$ have to hold.

sufficiently low (i.e., $G_i(\Psi_i)$ is large), his expected marginal incentives to intensify his manipulation and thereby to increase his posterior quality are close to $-\eta_L$. Therefore, he faces only weak manipulation incentives. He will even prefer to create no false information at all, if $\eta_L > 0$.

Contrary to the case of constant payoffs, the environment effect is now determined by the absolute probability of winning, $1-G_i(\Psi_i)$. The two updating weights, Σ_i and Σ_j, determine both the magnitude of Ψ_i and the shape of the cumulative distribution function $G_i(\cdot)$.

4.3. What kind of environment is susceptible to the manipulation of information?

Whether the payoff depends on the magnitude of the posterior lead or not in a given competition is often at the discretion of a superior organization. It is therefore an obvious question of interest, which form of competition leads to a higher manipulation intensity.

Recall that players' cost functions do not depend on the payoff structure. To analyze how the extent of information manipulation differs across environments in interior equilibria, it then suffices to compare the left-hand sides of (11) and (15). The manipulation intensity of player i will, thus, be higher under affine than under constant payoffs if and only if $\Delta_i > 0$ with Δ_i being defined as

$$\Delta_i = \eta_H - (\eta_H + \eta_L)\, G_i(\Psi_i) - (\bar{u}_H - \bar{u}_L)\, g_i(\Psi_i)\,.$$

The following proposition summarizes under which conditions this will be the case.

Proposition 2. *For interior equilibria, the following holds:*

(i) If $\eta_L > 0$, there exists a cutoff ξ such that Δ_i will be positive if and only if $\bar{q}_{i0} - \bar{q}_{j0} > \xi$.
(ii) If $\eta_L \leq 0$, there exist two intervals $(-\infty, \xi']$ and $[\xi'', \infty)$ with $\xi' \leq \xi''$ such that Δ_i will be positive if and only if $\bar{q}_{i0} - \bar{q}_{j0} \in (-\infty, \xi'] \cup [\xi'', \infty)$. Moreover, there exists a threshold $\bar{\eta}_L < 0$ such that $\xi' = \xi''$ for all $\eta_L < \bar{\eta}_L$.

The proposition shows that the prior lead is key to understand how manipulation intensities vary across payoff structures. Consider first the case of constant payoffs. This setting is particularly competitive if initial heterogeneity is small. In such a tight race, both players are willing to manipulate information. If initial heterogeneity becomes large, however, the winner is almost predetermined and players are not inclined to bear the cost of manipulating information. A large prior lead may even imply negligible information manipulation by both players. In contrast, with affine payoffs the competitiveness of the environment is much less dependent on the exact degree of heterogeneity. At least one player will be inclined to manipulate information in order to increase the posterior lead even if initial heterogeneity is large – if $\eta_L > 0$, one player wants to increase his lead, and if $\eta_L < 0$, both players compete for the margin of the lead.

As a consequence, affine payoffs generate larger manipulation intensities for unbalanced competitions than constant payoffs. In contrast, constant payoffs tend to generate higher manipulation intensities for balanced competitions. Only if η_L is smaller than zero and still sufficiently different from η_H, manipulation intensities under affine payoffs can dominate those under constant payoffs irrespective of the degree of heterogeneity. Intuitively, in this situation both players have incentives to invest in their perceived qualities under affine payoffs, and a sufficiently large value of $|\eta_L|$ ensures that also the incentives for the initially trailing player are substantial.

The findings above consider how manipulation intensities of single players differ in the two environments. As Proposition 2 shows, it might be the case that the manipulation intensity of one player is higher under affine payoffs compared to the case of constant payoffs while the opposite holds true for his opponent. In such cases, the effect of the payoff structure on the overall extent of information manipulation will depend on the individual weights for Bayesian updating, Σ_i, and initial quality uncertainty as measured by σ_{i0}^2. However, the proposition also implies clear-cut results. In particular, the sum of manipulation intensities will be larger under an affine payoff structure if $\eta_L \leq 0$ and initial player heterogeneity is large. In contrast, the sum of manipulation intensities will be larger under a constant payoff structure if $\eta_L \leq 0$ is intermediate and players are rather homogeneous when entering the game.

5. Institutions against information manipulation

Having established the equilibria for the two prominent classes of competitive environments, our setup allows to study how more rigorous institutions affect manipulation intensities. The specific payoff structures yield clear-cut results on the shape of the environment effect and therefore on the overall impact of the institutional setup on the players' inclination to produce false information.

5.1. The effect of institutions in an environment with constant payoffs

We start our analysis of the impact of institutions by first considering the environment with a constant payoff structure:

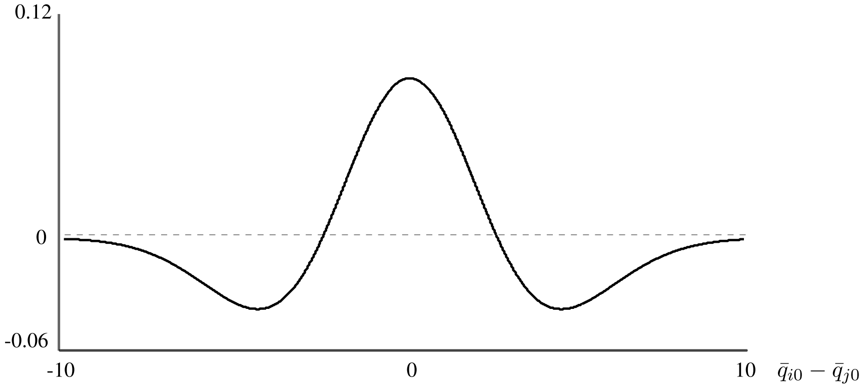

Fig. 1. The environment effect in dependence of $\bar{q}_{i0} - \bar{q}_{j0}$ for a constant payoff structure for $\sigma_{i0}^2 = \sigma_{j0}^2 = \sigma_{\varphi}^2 = 4$, $\beta = 0.5$, $\bar{u}_H = 6$, and $\bar{u}_L = 2$.

Proposition 3. *For an interior equilibrium, the following holds. If $u_H = \bar{u}_H$ and $u_L = \bar{u}_L$ with $\bar{u}_H > \bar{u}_L$, then the environment effect has the same sign for both players. For every $\beta > 0$, there exists a threshold $\chi_i^{const}(\beta) \geq 0$ such that manipulation intensity f_i^* will increase with more rigorous institutions if and only if $\left|\bar{q}_{i0} - \bar{q}_{j0}\right| > \chi_i^{const}(\beta)$.*

As described in Section 3, the information effect of institutions will be positive if and only if $\sigma_{i0}^2 > \beta^2\sigma_{\varphi}^2$. However, as Proposition 3 shows, under a constant payoff structure backfiring of more rigorous institutions can happen irrespective of the sign of the information effect if the prior lead is sufficiently large. In particular, the environment effect will be negative and dominates the information effect if the prior lead is sufficiently large.

To understand this finding, consider Fig. 1, which depicts the environment effect for different levels of players' initial heterogeneity, $\bar{q}_{i0} - \bar{q}_{j0}$. It illustrates that the environment effect will be negative for both players if and only if players' heterogeneity is large. The intuition is as follows. If one of the players has a substantial prior lead, the winner of the competition is almost predetermined. Consequently, both players' incentives to create false information are low. If in this situation β becomes smaller, however, the variance of the composed random variable δ_i will increase.[14] From the perspective of the players, the outcome of the competition will thus become less predictable. Therefore, the trailing player will have a real chance to win the competition, which restores both players' manipulation incentives. As a consequence, given a sufficiently large degree of initial heterogeneity, the environment effect of more rigorous institutions will induce both players to choose higher manipulation intensities in equilibrium.

At a cursory first glance, it seems counterintuitive that more rigorous institutions lead to an increase of the variance of δ (i.e., $\sigma_{\delta_i}^2$), which is crucial for the shape of the environment effect. On the one hand, a reduction in β indeed reduces the impact of false information that is not spread by the players, φ_i, as can be seen from (1). On the other hand, a reduction of β also induces the audience to rely more heavily on the received quality signals, which have become more trustworthy. In other words, the audience assigns larger weights Σ_A and Σ_B to the random variables φ_A, φ_B, q_A, and q_B when updating beliefs, which boosts the impact of chance. Expression (6) shows that the latter effect dominates the former, because $\partial\sigma_{\delta_i}^2/\partial\beta < 0$ as both Σ_i and Σ_j decrease with β.

Importantly, Proposition 3 shows that the environment effect will dominate the possibly positive information effect if initial heterogeneity is substantial. This result holds true even in the limit of very strong heterogeneity because the information effect converges to zero at a faster rate than the environment effect. Therefore, the shape of the overall effect of more thorough institutions resembles the shape of the environment effect: More rigorous institutions induce higher manipulation intensities by both players if and only if players are sufficiently heterogeneous. The intuition of this finding is as before. If payoffs are constant, a large prior lead induces low incentives to manipulate information, which can be restored by more trustworthy signals.

5.2. The effect of institutions in an environment with affine payoffs

Next, we study how institutions affect manipulation intensities under an affine payoff structure. Equation (15) shows that the information effect, described by $\beta\Sigma_i$ at the left-hand side, is identical to that in the environment with constant payoffs. The crucial difference to the constant payoff structure is the shape of the environment effect, which is depicted in Fig. 2. Under constant payoffs, the environment effect stems from a change in the players' marginal probability of winning the competition, $g_i(\Psi_i)$, which leads to the same impact of β on both players' manipulation intensities in equilibrium. Under affine payoffs, however, the environment effect stems from a change in the players' probability of becoming the competitor with a posterior lead. As player i's probability of becoming the competitor with a posterior lead will decrease if player j's

[14] See the definition of $\sigma_{\delta_i}^2$ in (6).

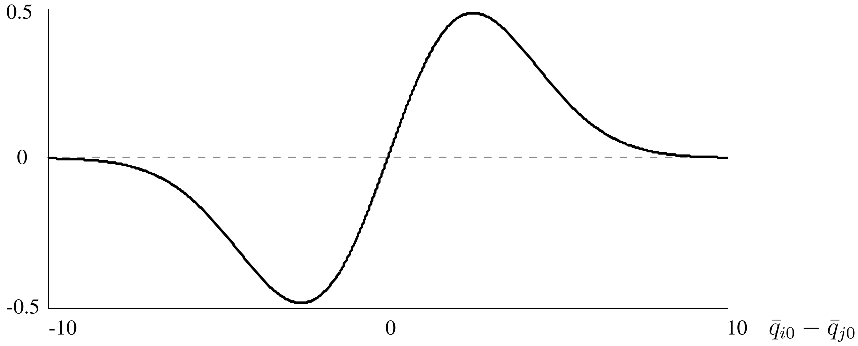

Fig. 2. The environment effect in dependence of $\bar{q}_{i0} - \bar{q}_{j0}$ for affine functions for $\sigma_{i0}^2 = \sigma_{j0}^2 = \sigma_{\varphi}^2 = 4$, $\beta = 0.5$, $\eta_H = 6$, and $\eta_L = 2$. The environment effect looks qualitatively identical for negative values of η_L.

respective probability increases and vice versa, β cannot have the same impact on both players' manipulation intensities (see Fig. 2).

Proposition 4. *For an interior equilibrium, the following holds. Suppose u_H and u_L are affine and given by (12) and (13) with $\eta_H > |\eta_L|$. Then, the environment effect is negative for player i if and only if $\bar{q}_{i0} < \bar{q}_{j0}$. Let $\sigma_{i0}^2 > \beta^2 \sigma_{\varphi}^2$, then there exists a threshold $\chi_i^{affine}(\beta) < 0$ such that player i will (weakly) increase his manipulation intensity as a response to more rigorous institutions if and only if $\bar{q}_{i0} - \bar{q}_{j0} < \chi_i^{affine}(\beta)$.*

Proposition 4 shows that, under affine payoffs, backfiring of more rigorous institutions can happen even if the information effect is strictly positive (i.e., $\sigma_{i0}^2 > \beta^2 \sigma_{\varphi}^2$).[15] In that situation, the corresponding environment effect of player i has to be negative and sufficiently strong, which is the case if $\bar{q}_{i0} < \bar{q}_{j0}$ and $|\bar{q}_{i0} - \bar{q}_{j0}|$ is sufficiently large.[16] The intuition for the relation $\bar{q}_{i0} < \bar{q}_{j0}$ is the following. Suppose player j enters the game with the prior lead $\bar{q}_{j0} - \bar{q}_{i0} > 0$. If in this situation institutions against information manipulation become more rigorous, the signals s_i and s_j will appear more trustworthy to the audience. Hence, by manipulating information, it will become easier for the trailing player i to catch up with j and to end up as the competitor with a posterior lead. As player i's posterior expected quality is particularly important to him if he becomes the player with a posterior lead because of $\eta_H > |\eta_L|$, his incentives to manipulate information become stronger. In contrast, more rigorous institutions deteriorate j's favorable starting position. His probability of becoming the competitor with a posterior lead decreases, which reduces his manipulation incentives.

6. Discussion

One important implication of our model is that more institutional scrutiny can lead to more information manipulation. In this section, we discuss to what extent this result depends on how we specify institutions in the model. The key premise that our results rest upon is that the audience will trust signals more intensely if institutions are more rigorous. In particular, the audience updates its priors about the players' qualities more strongly if institutions are more rigorous compared to a situation in which it receives the same signals but institutions are less rigorous. This idea is clearly not restricted to the exact specification of institutions that we use. However, it is also not necessarily true for every possible way how institutions could intervene in the information generating process. In the following, we will briefly discuss two alternative ways of modeling institutions in our setup.

6.1. Elimination of manipulation intensities

Instead of assuming that institutions affect all incorrect parts of the public signal, in this section we suppose that institutions can directly eliminate players' manipulation. Hence, institutions are able to discriminate between intentional information manipulation and exogenous false information. Technically, we thus eliminate the parameter β from our main model of Section 2, so that $\Sigma_i = \sigma_{i0}^2/(\sigma_{\varphi}^2 + \sigma_{i0}^2)$ $(i = A, B)$ and $\delta_i = \Sigma_i (q_i + \varphi_i) - \Sigma_j \left(q_j + \varphi_j\right)$ $(i, j = A, B;\ i \neq j)$, but Ψ_i, μ_{δ_i} and $\sigma_{\delta_i}^2$ are the same as defined in Section 3. We assume that the manipulation of player i's signal is detected with

[15] If, however, the information effect is negative because $\sigma_{i0}^2 < \beta^2 \sigma_{\varphi}^2$, it will reinforce the environment effect and more rigorous institutions will certainly backfire.

[16] The intuition for why $|\bar{q}_{i0} - \bar{q}_{j0}|$ needs to be sufficiently large is similar to that in the environment with constant payoffs. For a large prior lead, both effects tend to zero. However, the trailing player's incentives to invest in manipulation are restored if institutions are rigorous because this might give him the chance to win the competition. As a consequence, the environment effect converges to zero at a lower pace than the information effect.

probability $\gamma \cdot p(f_i)$ with $\gamma > 0$ denoting the degree of rigorousness by which institutions check information manipulation and $p' > 0$ indicating that the detection probability increases with player i's manipulation intensity. With probability $1 - \gamma \cdot p(f_i)$ the manipulation of signal s_i is not detected. In case of detection, manipulation intensity f_i is eliminated so that the audience receives the signal $s_i = q_i + \varphi_i$. The audience learns whether f_i has been eliminated or not. Solving the modified game leads to the following result:

Proposition 5. *In any interior equilibrium, the players will choose lower manipulation intensities if institutions become more rigorous.*

The proof of Proposition 5 shows that in our modified setting only the information effect exists and it has a clear sign. Irrespective of the payoff structure, the more rigorous the institutions the lower will be the equilibrium manipulation intensities of both players. Intuitively, how the audience updates its priors in response to given signals does not depend on the institutional scrutiny γ in this setup. If manipulation is removed, the audience takes the signal as unbiased; if it is not removed, the audience perceives the signal as biased. Thus, the updating weights do not depend on the degree of rigorousness, γ, and signals that are still manipulated do not seem more trustworthy to the audience if institutions are more rigorous. Players, therefore, react to more rigorous institutions by investing less in costly information manipulation. Two important assumptions drive this result: (i) institutions can and want to discriminate between intentional and unintentional spreading of false information, and (ii) players perfectly learn whether a piece of information has been investigated and debiased by the institutions. If these assumptions are not met, the audience would consider newly arriving information as more trustworthy if more rigorous institutions are in place. As a consequence, the same trade-off as in our main model would occur and, hence, more rigorous institutions may lead to higher manipulation intensities.

6.2. Sending additional signals

In this section, we take the stance that institutions do not directly influence the information provided by the players.[17] Rather institutions provide additional information on the players' qualities. This information materializes in a second public signal that cannot be distorted by the players:

$$s_{i2} = q_i + \varphi_{i2}.$$

This signal, s_{i2}, can be correlated with s_i, and we argue that institutions will be more rigorous if this correlation is low, i.e., if the information that institutions provide is less dependent on the information that the player spreads. If institutions are completely self-reliant, the two signals s_i and s_{i2} will be statistically independent. If, however, institutions are more lenient, they will replicate some of the information that is also entailed in the signal that is generated by the player, or they may copy from publicly available sources. Formally, we assume that φ_{i2} and φ_i are jointly normally distributed with correlation coefficient ρ, mean zero and variance one, where ρ is our measure for institutional scrutiny, i.e., the smaller ρ the more rigorous the institutions. For simplicity we further assume that $\sigma_{A0}^2 = \sigma_{B0}^2 = \sigma_0^2$. Considering interior equilibria, we obtain the following result:

Proposition 6. *More rigorous institutions lead to higher manipulation intensities, irrespective of whether the players' payoff structure is constant or affine.*

Proposition 6 shows that more rigorous institutions unambiguously aggravate information manipulation in this setup for both the case of constant and the case of affine payoffs. There are two reasons for this finding. First, irrespective of the competitive environment, more rigorous institutions induce the combination of public signals to be more informative. As a consequence, the audience adapts their priors more strongly to the newly arriving information. As the players anticipate this reaction of the audience, they have a stronger incentive to manipulate information at the beginning of the game. Second, the stronger the correlation among signals, the easier it is for the audience to detect manipulation. In the extreme case of $\rho = 1$, the audience can even fully infer the manipulation intensities from the difference of the signals. This second effect, reinforces the first one as, again, it implies that more rigorous institutions lead to higher manipulation intensities. Contrary to the main model, there is not even a countervailing effect of more rigorous institutions because the institutions do not eliminate the players' manipulation and, hence, do not reduce their expected returns from investing in information manipulation.

To sum up, the modifications of the main model have revealed two insights. First, if institutions perfectly discriminate between players' information manipulation and exogenous false information, more rigorous institutions will yield less information manipulation by the players because of reduced expected manipulation returns. Second, if institutions do neither eliminate endogenous nor exogenous information manipulation but improve the publicly available information and, hence, the trustworthiness of players' signals, more rigorous institutions aggravate information manipulation. Contrary to the modifications in this section, our main model comprises both of these effects. This explains why increasing institutional scrutiny in our main setup involves a trade-off and why the benefit thereof depends on the parameter constellation.

[17] Thus, again we eliminate the parameter β of the main model.

7. Applications

In this section, we suggest several applications for our information manipulation approach. For this purpose, we extend our one-stage model to a two-stage setting in which the additional second stage is used to further specify the respective application. We then show how simple micro foundations for the second stage yield insights on how to interpret our findings in various competitive situations. We will also show that the equilibrium payoffs of several widely used models for competition exactly coincide with the constant and affine payoffs that we use throughout the main part of this paper.

7.1. Elections and campaigning

One application of our model is in the realm of elections. Suppose that two candidates run for office in an election where the outcome only depends on their reputation (as, for example, in Majumdar and Mukand, 2004; Callander, 2008). Both candidates $i \in \{A, B\}$ with uncertain abilities q_i can try to influence the media via false information at the first stage, and compete at the second stage for votes. The electorate is heterogeneous in its tastes such that voter θ votes for candidate A if and only if $\bar{q}_{A1} > \bar{q}_{B1} + \theta$, where θ is distributed according to a cdf that is symmetric around 0. If the candidate that obtains more votes wins a high benefit $\bar{u}_H$ and the loser receives $\bar{u}_L$, this environment yields constant payoffs. However, in parliamentary elections, where also the election loser retains some say in the political process, the utility of winning typically depends on the posterior lead, which yields a payoff structure in which the posterior lead is important (for examples see Fishburn and Gehrlein, 1977; Grossman and Helpman, 1996; Grunewald et al., 2020a). As we show in the online appendix, affine payoffs also arise if politicians or lobbyists need to acquire capital for their campaign and then, at the second stage, invest in campaigning, which can be modeled as a Tullock contest or an all-pay auction.

7.2. Price competition

Our setup can also be applied to the competition among two firms A and B that may invest in deceptive advertising (e.g, Zinman and Zitzewitz, 2016) or online review manipulation (e.g, Dellarocas, 2006; Mayzlin et al., 2014) at stage one, and compete in prices at stage two. The audience is given by consumers who purchase a good from one of the two firms. Firm i $(i = A, B)$ offers a complex experience good i (e.g., a car, a computer, or a mobile phone) whose quality $q_i \sim N\left(\bar{q}_{i0}, \sigma_{i0}^2\right)$ is uncertain for the two firms and the consumers. At stage two, firms decide on prices. If consumers ignore product prices up to some bound and purchase on the basis of posterior expected quality only, the quality leader i serves the whole market and receives profits $\bar{u}_H > 0$, while firm $j \neq i$ with the lower expected quality earns profits $\bar{u}_L = 0$. Hence, we obtain constant payoffs. In contrast, as we show in the online appendix, firms anticipate an affine payoff structure if stage two is described by a simplified version of the price competition game with vertical product differentiation considered by Shaked and Sutton (1982).

7.3. Influence activities on the job

As a final application, consider the case of influence activities and job promotions in firms. Following the papers by Milgrom (1988) and Milgrom and Roberts (1988), in some situations politicking by workers might be important to get promoted to a better job in the hierarchy. Suppose that two workers A and B try to influence their superior that is in charge of the promotion decision by creating false information within the workforce. If the worker with the better perceived ability $\bar{q}_{i1}$ is promoted and earns high income $\bar{u}_H$, whereas his opponent stays at his current job and receives the lower income $\bar{u}_L$, we will end up with constant payoffs. If, however, the wage increase upon promotion depends on how strongly the winner of the career competition has outperformed his opponent, the given scenario will be better captured by an affine payoff structure.

8. Conclusion

This paper provides a general framework for investigating how the incentives to manipulate information depend on the competitive environment and on the institutions that are in place against information manipulation. For this purpose, we consider a model in which two players compete against each other by creating false information about their expected qualities. We derive two main findings: First, if the magnitude of the posterior lead is relevant for the players' payoffs, at least one player will spread more false information compared to the case with constant payoffs if and only if the players are ex ante sufficiently heterogeneous. Second, more rigorous institutions against information manipulation induce people to rely more heavily on new information, which can increase manipulation intensities in equilibrium.

Appendix A

Proof of Proposition 1. In general, the optimal extent of manipulation can be interior, $f_i^* > 0$, as well as a corner solution, $f_i^* = 0$. In any interior equilibrium in pure strategies, f_i^* $(i = A, B)$ is described by the players' first-order conditions. Applying Leibniz's formula, the first-order condition of player i can be computed as

$$\beta\Sigma_i \cdot u_H(0) \cdot g_i\left(\Psi_i + \beta\Sigma_j(\tilde{f}_j - \hat{f}_j) - \beta\Sigma_i(f_i - \hat{f}_i)\right)$$
$$+ \beta\Sigma_i \cdot \int_{\Psi_i+\beta\Sigma_j(\tilde{f}_j-\hat{f}_j)-\beta\Sigma_i(f_i-\hat{f}_i)}^{\infty} u_H'\left(\delta_i - \Psi_i - \beta\Sigma_j(\tilde{f}_j - \hat{f}_j) + \beta\Sigma_i(f_i - \hat{f}_i)\right) g_i(\delta_i)\, d\delta_i$$
$$- \beta\Sigma_i \cdot u_L(0) \cdot g_i\left(\Psi_i + \beta\Sigma_j(\tilde{f}_j - \hat{f}_j) - \beta\Sigma_i(f_i - \hat{f}_i)\right)$$
$$- \beta\Sigma_i \cdot \int_{-\infty}^{\Psi_i+\beta\Sigma_j(\tilde{f}_j-\hat{f}_j)-\beta\Sigma_i(f_i-\hat{f}_i)} u_L'\left(\Psi_i + \beta\Sigma_j(\tilde{f}_j - \hat{f}_j) - \beta\Sigma_i(f_i - \hat{f}_i) - \delta_i\right) g_i(\delta_i)\, d\delta_i$$
$$= c'(f_i).$$

Using the fact that beliefs are derived from the players' actual manipulation intensities, i.e., $\tilde{f}_i = \hat{f}_i = f_i^*$ $(i = A, B)$, in any interior equilibrium we obtain the result summarized in the proposition. □

Proof of Corollary 1. Condition (10) is a sufficient condition to ensure strict concavity of the players' objective functions. The left-hand side of (10) uses the fact that $-g_i'$ attains its maximum in the right inflection point at $\Sigma_i\bar{q}_{i0} - \Sigma_j\bar{q}_{j0} + \sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}$. Combined with the assumption that marginal costs become infinitely large for high manipulation intensities, (10) guarantees existence and uniqueness of a pure-strategy equilibrium. Equation (11) is obtained from (7) by noting that both integrals are zero under a constant payoff structure, and by replacing $u_H(0)$ with $\bar{u}_H$ and $u_L(0)$ with $\bar{u}_L$. □

Proof of Corollary 2. Inequality (14) describes a sufficient condition for strict concavity of i's objective function. Together with the assumption that $c'(f_i) \to \infty$ for $f_i \to \infty$ it therefore guarantees existence and uniqueness of a pure-strategy equilibrium. It is obtained from i's second-order condition assuming that it holds in the most restrictive case. Here, we use the fact that the density g_i attains its maximum at μ_{δ_i}.

Player i will choose the corner solution $f_i^* = 0$ if the derivative of his objective function is negative at $f_i = 0$. From (15) we know that this is the case exactly if

$$\eta_H - (\eta_H + \eta_L)\, G_i(\Psi_i) < 0. \tag{A.16}$$

As assumption $\eta_H \geq |\eta_L|$ implies that $\eta_H + \eta_L \geq 0$, condition (A.16) may hold or not. Hence, only if $(\eta_H + \eta_L)\, G_i(\Psi_i)$ exceeds η_H, i.e., if Ψ_i becomes large, inequality (A.16) will be satisfied. Note that Ψ_i can indeed become arbitrarily large, for example if $\bar{q}_{j0} > \bar{q}_{i0}$ and both means substantially differ. If the condition for a corner solution does not hold, an interior solution will exist, being described by equation (15). □

Proof of Proposition 2. For part (i) assume that $\eta_L > 0$. We have

$$\Delta_i = \eta_H - (\eta_H + \eta_L)\, G_i(\Psi_i) - (\bar{u}_H - \bar{u}_L)\, g_i(\Psi_i)$$
$$= \eta_H - (\eta_H + \eta_L)\, \Phi\left(\frac{-(\bar{q}_{i0} - \bar{q}_{j0})}{\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}}\right) - \frac{\bar{u}_H - \bar{u}_L}{\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}}\phi\left(\frac{-(\bar{q}_{i0} - \bar{q}_{j0})}{\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}}\right),$$

with Φ denoting the cumulative distribution function and ϕ the density of the standard normal distribution. The derivative with respect to $\bar{q}_{i0} - \bar{q}_{j0}$ yields

$$\frac{\eta_H + \eta_L}{\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}}\phi\left(\frac{-(\bar{q}_{i0} - \bar{q}_{j0})}{\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}}\right) + \frac{\bar{u}_H - \bar{u}_L}{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}\phi'\left(\frac{-(\bar{q}_{i0} - \bar{q}_{j0})}{\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}}\right).$$

By using the property $\phi'(x) = -x\phi(x)$ of the standard normal density, the derivative will be positive iff

$$\frac{\eta_H + \eta_L}{\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}} + \frac{\bar{u}_H - \bar{u}_L}{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}\frac{\bar{q}_{i0} - \bar{q}_{j0}}{\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}} > 0 \Leftrightarrow$$
$$\eta_H + \eta_L + \frac{\bar{u}_H - \bar{u}_L}{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}\left(\bar{q}_{i0} - \bar{q}_{j0}\right) > 0. \tag{A.17}$$

We conclude that the derivative will be positive if and only if $\bar{q}_{i0} - \bar{q}_{j0}$ is sufficiently large. Hence, Δ_i will be monotonically increasing starting from one point onwards and will be negative before that point. Therefore, there can exist at most one ξ such that Δ_i is positive if and only if $\bar{q}_{i0} - \bar{q}_{j0} > \xi$. To show that ξ exists, it suffices to show that $\lim_{\bar{q}_{i0}-\bar{q}_{j0}\to\infty} \Delta_i > 0$, which is obviously true.

For (ii) suppose that $\eta_L < 0$. For $|\bar{q}_{i0} - \bar{q}_{j0}| \to \infty$, the manipulation intensity of i will converge to zero under constant payoffs and to some positive value under affine payoffs. However, there will nevertheless be an interval of values for $\bar{q}_{i0} - \bar{q}_{j0}$ such that i's manipulation intensity is higher under constant payoffs. To see this, we make use of (A.17), which implies that the minimum of Δ_i will be given by

$$\eta_H - (\eta_H + \eta_L)\, \Phi\left(\frac{\eta_H + \eta_L}{\bar{u}_H - \bar{u}_L}\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}\right) - \frac{\bar{u}_H - \bar{u}_L}{\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}}\phi\left(\frac{\eta_H + \eta_L}{\bar{u}_H - \bar{u}_L}\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}\right). \tag{A.18}$$

For $\eta_L = 0$ this term is

$$\eta_H\left[1 - \Phi\left(\frac{\eta_H}{\bar{u}_H - \bar{u}_L}\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}\right)\right] - \frac{\bar{u}_H - \bar{u}_L}{\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}}\phi\left(\frac{\eta_H}{\bar{u}_H - \bar{u}_L}\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}\right) < 0$$

$$\Leftrightarrow \frac{\eta_H}{\bar{u}_H - \bar{u}_L}\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}\left[1 - \Phi\left(\frac{\eta_H}{\bar{u}_H - \bar{u}_L}\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}\right)\right] - \phi\left(\frac{\eta_H}{\bar{u}_H - \bar{u}_L}\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}\right) < 0.$$

We know that this inequality is fulfilled since the hazard rate of the standard normal distribution is larger than its argument at any positive argument (see, for example, Baricz, 2008). From (A.17), we know that Δ_i decreases until it reaches its minimum and increases afterward. With $\lim_{|\bar{q}_{i0}-\bar{q}_{j0}|\to\infty} \Delta_i > 0$, it is therefore clear that there exist two thresholds ξ' and ξ'' such that Δ_i is positive if and only if $\bar{q}_{i0} - \bar{q}_{j0} \in (-\infty, \xi'] \cup [\xi'', \infty)$, which establishes the first part of result (ii) for values of η_L close to zero.

Moreover, the minimum of Δ_i in (A.18) is differentiable in η_L with derivative

$$- \Phi\left(\frac{\eta_H + \eta_L}{\bar{u}_H - \bar{u}_L}\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}\right) - (\eta_H + \eta_L)\,\phi\left(\frac{\eta_H + \eta_L}{\bar{u}_H - \bar{u}_L}\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}\right)\frac{\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}}{\bar{u}_H - \bar{u}_L}$$
$$- \phi'\left(\frac{\eta_H + \eta_L}{\bar{u}_H - \bar{u}_L}\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}\right).$$

By using the property $\phi'(x) = -x\phi(x)$, the derivative is given by

$$- \Phi\left(\frac{\eta_H + \eta_L}{u_H - u_L}\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}\right),$$

which is negative. Hence, there will exist two thresholds ξ' and ξ'' such that Δ_i is positive if and only if $\bar{q}_{i0} - \bar{q}_{j0} \in (-\infty, \xi'] \cup [\xi'', \infty)$ for all values of η_L larger than some negative threshold $\bar{\eta}_L$, which establishes (ii).

Finally, we can state a condition for which $\bar{\eta}_L$ will be larger than the smallest possible value of η_L which is, by assumption, given by $-\eta_H$, or whether the two intervals $(-\infty, \xi']$ and $[\xi'', \infty)$ are always disjoint. The threshold $\bar{\eta}_L$ will be larger than $-\eta_H$ if and only if the minimum of Δ_i (given by (A.18)) evaluated at $\eta_L = -\eta_H$ is positive:

$$\eta_H - \frac{\bar{u}_H - \bar{u}_L}{\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}}\phi(0) > 0 \Leftrightarrow \eta_H \geq (\bar{u}_H - \bar{u}_L)g_i\left(\Sigma_i\bar{q}_{i0} - \Sigma_j\bar{q}_{j0}\right).$$

If this is the case, the minimum of Δ_i is positive for all $\eta_L < \bar{\eta}_L$, and for all $\eta_L \in [\bar{\eta}_L, 0)$ the minimum will be negative. □

Proof of Proposition 3. By transforming the normal distribution into the standard normal distribution with density ϕ, equation (11) can be rewritten as

$$\frac{\beta\Sigma_i(\bar{u}_H - \bar{u}_L)}{\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}}\phi\left(\frac{\bar{q}_{j0} - \bar{q}_{i0}}{\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}}\right) = c'\left(f_i^*\right). \tag{A.19}$$

Computing the derivative of the left-hand side with respect to β yields

$$\frac{(\bar{u}_H - \bar{u}_L)\sigma_{i0}^2}{\left(\beta^2\sigma_\varphi^2 + \sigma_{i0}^2\right)^2}\frac{\sigma_{i0}^2 - \beta^2\sigma_\varphi^2}{\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}}\phi\left(\frac{\bar{q}_{j0} - \bar{q}_{i0}}{\sqrt{\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2}}\right) + \frac{(\bar{u}_H - \bar{u}_L)\beta\sigma_{i0}^2}{\beta^2\sigma_\varphi^2 + \sigma_{i0}^2}\frac{\left(\Sigma_i^2 + \Sigma_j^2\right)\beta\sigma_\varphi^2}{\left(\Sigma_i\sigma_{i0}^2 + \Sigma_j\sigma_{j0}^2\right)^{\frac{3}{2}}}\times$$

$$\left[\phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{\sqrt{\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2}}\right)+\frac{\bar{q}_{j0}-\bar{q}_{i0}}{\sqrt{\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2}}\phi'\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{\sqrt{\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2}}\right)\right].$$

Using the fact that $\phi'(x)=-x\phi(x)$, the derivative will be positive iff

$$\frac{\sigma_{i0}^2-\beta^2\sigma_\varphi^2}{\beta^2\sigma_\varphi^2+\sigma_{i0}^2}+\frac{\beta^2\sigma_\varphi^2\left(\Sigma_i^2+\Sigma_j^2\right)}{\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2}\cdot\left[1-\frac{\left(\bar{q}_{j0}-\bar{q}_{i0}\right)^2}{\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2}\right]>0.$$

Whereas the first addend corresponds to the information effect (see (9)), the second addend describes the environment effect. As this expression is identical for both players, the environment effect has the same sign for A and B. Moreover, it is monotonically decreasing with the degree of initial player heterogeneity as measured by $\left|\bar{q}_{i0}-\bar{q}_{j0}\right|$. If its value becomes sufficiently large, the whole derivative will be negative. □

Proof of Proposition 4. As δ_i is normally distributed with mean $\Sigma_i\bar{q}_{i0}-\Sigma_j\bar{q}_{j0}$ and variance $\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2$, equation (15) can be rewritten as

$$\beta\Sigma_i\left[\eta_H-(\eta_H+\eta_L)\,\Phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{\sqrt{\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2}}\right)\right]=c'\left(f_i^*\right). \tag{A.20}$$

The derivative with respect to β of the left-hand side will be positive iff

$$\begin{aligned}
&\left[\Sigma_i-\frac{2\beta^2\sigma_{i0}^2\sigma_\varphi^2}{\left(\beta^2\sigma_\varphi^2+\sigma_{i0}^2\right)^2}\right]\left[\eta_H-(\eta_H+\eta_L)\,\Phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{\sqrt{\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2}}\right)\right]\\
&+(\eta_H+\eta_L)\,\beta\Sigma_i\phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{\sqrt{\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2}}\right)\frac{\bar{q}_{i0}-\bar{q}_{j0}}{\left(\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2\right)^{\frac{3}{2}}}\left(\Sigma_i^2+\Sigma_j^2\right)\beta\sigma_\varphi^2>0\\
\Leftrightarrow&\frac{\sigma_{i0}^2-\beta^2\sigma_\varphi^2}{\beta^2\sigma_\varphi^2+\sigma_{i0}^2}\left[\eta_H-(\eta_H+\eta_L)\,\Phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{\sqrt{\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2}}\right)\right]\\
&+(\eta_H+\eta_L)\,\beta\phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{\sqrt{\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2}}\right)\frac{\left(\bar{q}_{i0}-\bar{q}_{j0}\right)\beta\sigma_\varphi^2}{\left(\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2\right)^{\frac{3}{2}}}\left(\Sigma_i^2+\Sigma_j^2\right)>0.
\end{aligned} \tag{A.21}$$

We are looking for values of initial heterogeneity as measured by $\bar{q}_{i0}-\bar{q}_{j0}$ such that this inequality is fulfilled. First, note that for $\bar{q}_{i0}-\bar{q}_{j0}>0$ the environment effect of player i (i.e., the second line of (A.21)) is positive. Moreover, in the first line of inequality (A.21), the first term is positive by assumption and the second term is positive in any interior equilibrium so that the information effect is positive as well. Hence, the inequality is fulfilled. Suppose now $\bar{q}_{i0}-\bar{q}_{j0}<0$. In this case, the environment effect of player i is strictly negative. To show the existence of $\chi_i^{\text{affine}}(\beta)$, we follow two steps.

Step 1: The environment effect can dominate the information effect
Suppose $\eta_L>0$. Consider initial heterogeneity $\bar{q}_{i0}-\bar{q}_{j0}<0$ small enough such that $\Phi(\cdot)$ is almost equal to $\frac{\eta_H}{\eta_H+\eta_L}$. In this case, the information effect becomes arbitrarily small while the environment effect remains strictly negative. Hence, the effect of β on manipulation intensity will be strictly negative. Now, suppose $\eta_L<0$. If $\bar{q}_{i0}-\bar{q}_{j0}\to-\infty$, the information effect converges to $\frac{\sigma_{i0}^2-\beta^2\sigma_\varphi^2}{\beta^2\sigma_\varphi^2+\sigma_{i0}^2}\eta_H>0$, whereas the environment effect converges to $-\infty$ so that the overall effect of β on manipulation intensity is strictly negative. □

Step 2: The environment effect has only one root in $\bar{q}_{i0}-\bar{q}_{j0}$
To prove this claim consider the derivative of the left-hand side of (A.21) with respect to $\bar{q}_{i0}-\bar{q}_{j0}$:

$$\frac{\sigma_{i0}^2-\beta^2\sigma_\varphi^2}{\beta^2\sigma_\varphi^2+\sigma_{i0}^2}\frac{\eta_H+\eta_L}{\sqrt{\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2}}\phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{\sqrt{\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2}}\right)$$

$$-\left(\eta_H+\eta_L\right)\beta\phi'\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{\sqrt{\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2}}\right)\frac{\left(\bar{q}_{i0}-\bar{q}_{j0}\right)\beta\sigma_\varphi^2}{\left(\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2\right)^2}\left(\Sigma_i^2+\Sigma_j^2\right)$$

$$+\left(\eta_H+\eta_L\right)\beta\phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{\sqrt{\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2}}\right)\frac{\beta\sigma_\varphi^2}{\left(\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2\right)^{\frac{3}{2}}}\left(\Sigma_i^2+\Sigma_j^2\right).$$

Using that $\phi'(x)=-x\phi(x)$, this expression will be positive if and only if

$$\frac{\sigma_{i0}^2-\beta^2\sigma_\varphi^2}{\beta\left(\Sigma_i^2+\Sigma_j^2\right)\left(\beta^2\sigma_\varphi^2+\sigma_{i0}^2\right)}-\frac{\left(\bar{q}_{i0}-\bar{q}_{j0}\right)^2\beta\sigma_\varphi^2}{\left(\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2\right)^2}+\frac{\beta\sigma_\varphi^2}{\Sigma_i\sigma_{i0}^2+\Sigma_j\sigma_{j0}^2}>0.$$

Hence, the derivative is positive if and only if $|\bar{q}_{i0}-\bar{q}_{j0}|$ is small. The result from step 1 further implies that the overall effect of β on manipulation intensity converges to a negative value for $\bar{q}_{i0}-\bar{q}_{j0}$ sufficiently small. Moreover, the effect is positive at $\bar{q}_{i0}-\bar{q}_{j0}=0$ and remains positive for larger differences. Summing up, the effect is negative at some negative values for $\bar{q}_{i0}-\bar{q}_{j0}$, it increases monotonically in an interval around 0 (and decreases everywhere else), becomes positive and remains positive thereafter. Hence, there can only be one root. □

Taking steps 1 and 2 together implies the existence of a threshold $\chi_i^{\text{affine}}(\beta)$ such that player i will (weakly) increase his manipulation intensity as a response to more rigorous institutions – i.e., $\partial f_i^*/\partial\beta\leq 0$ – if and only if $\bar{q}_{i0}-\bar{q}_{j0}<\chi_i^{\text{affine}}(\beta)$. □

Proof of Proposition 5. Player i maximizes

$$\gamma p\left(f_i\right)\gamma p\left(\tilde{f}_j\right)\int_{\Psi_i}^{\infty}u_H\left(\delta_i-\Psi_i\right)g_i\left(\delta_i\right)d\delta+\gamma p\left(f_i\right)\gamma p\left(\tilde{f}_j\right)\int_{-\infty}^{\Psi_i}u_L\left(\Psi_i-\delta_i\right)g_i\left(\delta_i\right)d\delta_i$$

$$+\gamma p\left(f_i\right)\left(1-\gamma p\left(\tilde{f}_j\right)\right)\int_{\Psi_i+\Sigma_j\left(\tilde{f}_j-\hat{f}_j\right)}^{\infty}u_H\left(\delta_i-\Psi_i-\Sigma_j\left(\tilde{f}_j-\hat{f}_j\right)\right)g_i\left(\delta_i\right)d\delta_i$$

$$+\gamma p\left(f_i\right)\left(1-\gamma p\left(\tilde{f}_j\right)\right)\int_{-\infty}^{\Psi_i+\Sigma_j\left(\tilde{f}_j-\hat{f}_j\right)}u_L\left(\Psi_i+\Sigma_j\left(\tilde{f}_j-\hat{f}_j\right)-\delta_i\right)g_i\left(\delta_i\right)d\delta_i$$

$$+\left(1-\gamma p\left(f_i\right)\right)\gamma p\left(\tilde{f}_j\right)\int_{\Psi_i-\Sigma_i\left(f_i-\hat{f}_i\right)}^{\infty}u_H\left(\delta_i-\Psi_i+\Sigma_i\left(f_i-\hat{f}_i\right)\right)g_i\left(\delta_i\right)d\delta_i$$

$$+\left(1-\gamma p\left(f_i\right)\right)\gamma p\left(\tilde{f}_j\right)\int_{-\infty}^{\Psi_i-\Sigma_i\left(f_i-\hat{f}_i\right)}u_L\left(\Psi_i-\Sigma_i\left(f_i-\hat{f}_i\right)-\delta_i\right)g_i\left(\delta_i\right)d\delta_i$$

$$+\left(1-\gamma p\left(f_i\right)\right)\left(1-\gamma p\left(\tilde{f}_j\right)\right)\int_{\Psi_i+F}^{\infty}u_H\left(\delta_i-\Psi_i-F\right)g_i\left(\delta_i\right)d\delta_i$$

$$+\left(1-\gamma p\left(f_i\right)\right)\left(1-\gamma p\left(\tilde{f}_j\right)\right)\int_{-\infty}^{\Psi_i+F}u_L\left(\Psi_i+F-\delta_i\right)g_i\left(\delta_i\right)d\delta_i-c\left(f_i^*\right)$$

with $F:=\Sigma_j\left(\tilde{f}_j-\hat{f}_j\right)-\Sigma_i\left(f_i-\hat{f}_i\right)$. In an interior equilibrium, the first-order condition holds. Moreover, in equilibrium, beliefs and chosen manipulation intensities coincide (i.e., $\tilde{f}_i=\hat{f}_i=f_i^*$), implying

$$\Sigma_i\left(1-\gamma p\left(f_i^*\right)\right)\left[\left(u_H(0)-u_L(0)\right)g_i\left(\Psi_i\right)+\int_{\Psi_i}^{\infty}u_H'\left(\delta_i-\Psi_i\right)g_i\left(\delta_i\right)d\delta_i-\int_{-\infty}^{\Psi_i}u_L'\left(\Psi_i-\delta_i\right)g_i\left(\delta_i\right)d\delta_i\right]=c'\left(f_i^*\right).$$

Differentiating f_i^* implicitly with respect to γ yields

$$\frac{df_i^*}{d\gamma} = \frac{\Sigma_i p\left(f_i^*\right) \cdot A}{-\Sigma_i \gamma p'\left(f_i^*\right) \cdot A - c''\left(f_i^*\right)} < 0$$

with

$$A := \left(u_H(0) - u_L(0)\right) g_i(\Psi_i) + \int_{\Psi_i}^{\infty} u_H'(\delta_i - \Psi_i)\, g_i(\delta_i)\, d\delta_i - \int_{-\infty}^{\Psi_i} u_L'(\Psi_i - \delta_i)\, g_i(\delta_i)\, d\delta_i,$$

which is positive as we consider interior equilibria. □

Proof of Proposition 6. Let V denote the covariance matrix such that

$$\begin{bmatrix} \varphi_i \\ \varphi_{i2} \\ q_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ \bar{q}_{i0} \end{bmatrix}, V\right) \qquad \text{with} \qquad V = \begin{bmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & \sigma_0^2 \end{bmatrix}.$$

Note that

$$\begin{bmatrix} q_i + f_i + \varphi_i \\ q_i + \varphi_{i2} \\ q_i \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} \varphi_i \\ \varphi_{i2} \\ q_i \end{bmatrix} + \boldsymbol{b} \qquad \text{with} \qquad \boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{b} = \begin{bmatrix} f_i \\ 0 \\ 0 \end{bmatrix}.$$

As in Holmström (1999), the audience has a point belief about manipulation intensity f_i, denoted by $\hat{f}_i$. Hence, its belief about the joint distribution of the two signals and player i's quality is given by

$$\begin{bmatrix} q_i + f_i + \varphi_i \\ q_i + \varphi_{i2} \\ q_i \end{bmatrix} \sim N\left(\begin{bmatrix} \bar{q}_{i0} + \hat{f}_i \\ \bar{q}_{i0} \\ \bar{q}_{i0} \end{bmatrix}, \boldsymbol{V}\right) \quad \text{with} \qquad \boldsymbol{V} = \boldsymbol{A} V \boldsymbol{A}' = \begin{bmatrix} 1+\sigma_0^2 & \rho+\sigma_0^2 & \sigma_0^2 \\ \rho+\sigma_0^2 & 1+\sigma_0^2 & \sigma_0^2 \\ \sigma_0^2 & \sigma_0^2 & \sigma_0^2 \end{bmatrix}.$$

Following DeGroot (1970), the audience's belief about player i's quality after observing the signals s_i and s_{i2} is given by

$$\begin{aligned} E\left[q_i | s_i, s_{i2}\right] &= \bar{q}_{i0} + [\sigma_0^2, \sigma_0^2] \begin{bmatrix} 1+\sigma_0^2 & \rho+\sigma_0^2 \\ \rho+\sigma_0^2 & 1+\sigma_0^2 \end{bmatrix}^{-1} \begin{bmatrix} s_i - \bar{q}_{i0} - \hat{f}_i \\ s_{i2} - \bar{q}_{i0} \end{bmatrix} \\ &= \bar{q}_{i0} + \frac{1}{(1+\sigma_0^2)^2 - (\rho+\sigma_0^2)^2} [\sigma_0^2, \sigma_0^2] \begin{bmatrix} 1+\sigma_0^2 & -(\rho+\sigma_0^2) \\ -(\rho+\sigma_0^2) & 1+\sigma_0^2 \end{bmatrix} \begin{bmatrix} s_i - \bar{q}_{i0} - \hat{f}_i \\ s_{i2} - \bar{q}_{i0} \end{bmatrix} \\ &= \bar{q}_{i0} + \frac{1}{1+2\sigma_0^2+\rho} [\sigma_0^2, \sigma_0^2] \begin{bmatrix} s_i - \bar{q}_{i0} - \hat{f}_i \\ s_{i2} - \bar{q}_{i0} \end{bmatrix} \\ &= \bar{q}_{i0} + \frac{\sigma_0^2}{1+2\sigma_0^2+\rho} [s_i - \bar{q}_{i0} - \hat{f}_i + s_{i2} - \bar{q}_{i0}] =: \bar{q}_{i1}. \end{aligned}$$

Define

$$\Sigma := \frac{\sigma_0^2}{1+2\sigma_0^2+\rho} \tag{A.22}$$

and $\bar{\varphi}_i := \varphi_i + \varphi_{i2}$ such that $\bar{q}_{i1} - \bar{q}_{j1} = \delta_i - \Psi_i - \Sigma(\tilde{f}_j - \hat{f}_j) + \Sigma(f_i - \hat{f}_i)$, with δ_i being stochastic and Ψ_i embracing the exogenous deterministic elements:

$$\delta_i := \Sigma\left(2q_i + \bar{\varphi}_i - 2q_j - \bar{\varphi}_j\right) \qquad \Psi_i := (1-2\Sigma)\left(\bar{q}_{j0} - \bar{q}_{i0}\right). \tag{A.23}$$

Since any convolution of two normal densities again yields a normal density (e.g., Ross, 2010, pp. 35, 67–68), the composed random variable δ_i is normally distributed: $\delta_i \sim N\left(\mu_{\delta_i}, \sigma_{\delta_i}^2\right)$ with

$$\mu_{\delta_i} := 2\Sigma(\bar{q}_{i0} - \bar{q}_{j0}) \tag{A.24}$$

$$\text{and} \quad \sigma_{\delta_i}^2 := 8\Sigma^2\sigma_0^2 + 4\Sigma^2(1+\rho) = 4\Sigma^2(2\sigma_0^2+1+\rho) = 4\Sigma\sigma_0^2. \tag{A.25}$$

Let g_i denote the density of δ_i, and G_i the corresponding cumulative distribution function. Consequently, player i's objective function (3) can be rewritten as

$$\int_{\Psi_i+\Sigma(\tilde{f}_j-\hat{f}_j)-\Sigma(f_i-\hat{f}_i)}^{\infty} u_H\left(\delta_i-\Psi_i-\Sigma(\tilde{f}_j-\hat{f}_j)+\Sigma(f_i-\hat{f}_i)\right)g_i(\delta_i)\,d\delta_i$$
$$+\int_{-\infty}^{\Psi_i+\Sigma(\tilde{f}_j-\hat{f}_j)-\Sigma(f_i-\hat{f}_i)} u_L\left(\Psi_i+\Sigma(\tilde{f}_j-\hat{f}_j)-\Sigma(f_i-\hat{f}_i)-\delta_i\right)g_i(\delta_i)\,d\delta_i-c(f_i).$$

Hence, the first-order condition is given by

$$\Sigma\left[(u_H(0)-u_L(0))\,g_i(\Psi_i)+\int_{\Psi_i}^{\infty}u_H'(\delta_i-\Psi_i)\,g_i(\delta_i)\,d\delta_i-\int_{-\infty}^{\Psi_i}u_L'(\Psi_i-\delta_i)g_i(\delta_i)\,d\delta_i\right]=c'\left(f_i^*\right).$$

Now, we can check whether backfiring of more rigorous institutions can also occur in our modified setting. We start with the case of constant payoffs. Equation (11) becomes

$$\Sigma(\bar{u}_H-\bar{u}_L)g_i(\Psi)=c'(f_i^*).$$

By transforming the normal distribution into the standard normal distribution with density ϕ, we get

$$\frac{\sqrt{\Sigma}(\bar{u}_H-\bar{u}_L)}{2\sigma_0}\phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\sqrt{\Sigma}}\right)=c'\left(f_i^*\right). \tag{A.26}$$

Computing the derivative of the left-hand side with respect to ρ yields

$$\frac{-\sigma_0^2}{(1+2\sigma_0+\rho)^2}\frac{\bar{u}_H-\bar{u}_L}{4\sigma_0}\left[\frac{1}{\sqrt{\Sigma}}\phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\sqrt{\Sigma}}\right)-\phi'\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\sqrt{\Sigma}}\right)\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\Sigma}\right].$$

Using the fact that $\phi'(x)=-x\phi(x)$, the derivative will be negative iff

$$\frac{-\sigma_0^2}{(1+2\sigma_0+\rho)^2}\frac{\bar{u}_H-\bar{u}_L}{4\sigma_0}\phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\sqrt{\Sigma}}\right)\frac{1}{\sqrt{\Sigma}}\left[1+\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\sqrt{\Sigma}}\right)^2\right]<0,$$

which is true. Hence, the stronger the correlation (i.e., the worse institutions) the lower will be player i's incentive to manipulate.

Now, we turn to the case of an affine payoff structure. In our modified setting, (15) becomes

$$\Sigma[\eta_H-(\eta_H+\eta_L)G_i(\Psi_i)]=c'(f_i^*).$$

The left-hand side can be rewritten as

$$\Sigma\left[\eta_H-(\eta_H+\eta_L)\Phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\sqrt{\Sigma}}\right)\right].$$

The derivative with respect to ρ will be negative iff

$$\frac{-\sigma_0^2}{(1+2\sigma_0+\rho)^2}\left[\eta_H-(\eta_H+\eta_L)\Phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\sqrt{\Sigma}}\right)+\Sigma(\eta_H+\eta_L)\phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\sqrt{\Sigma}}\right)\frac{1}{2}\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\Sigma^{\frac{3}{2}}}\right]<0$$

$$\Leftrightarrow \eta_H-(\eta_H+\eta_L)\Phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\sqrt{\Sigma}}\right)+(\eta_H+\eta_L)\phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\sqrt{\Sigma}}\right)\frac{\bar{q}_{j0}-\bar{q}_{i0}}{4\sigma_0\sqrt{\Sigma}}>0. \tag{A.27}$$

As we consider interior equilibria, $\eta_H-(\eta_H+\eta_L)\,\Phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\sqrt{\Sigma}}\right)>0$ must hold, so that inequality (A.27) is satisfied for all $\bar{q}_{j0}\geq\bar{q}_{i0}$. Now, suppose that $\bar{q}_{j0}<\bar{q}_{i0}$. The derivative of the left-hand side of (A.27) with respect to $\bar{q}_{j0}-\bar{q}_{i0}$ yields

$$\frac{\eta_H+\eta_L}{4\sigma_0\sqrt{\Sigma}}\left[\phi'\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\sqrt{\Sigma}}\right)\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\sqrt{\Sigma}}-\phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\sqrt{\Sigma}}\right)\right].$$

By using $\phi'(x)=-x\phi(x)$, this expression can be rewritten as

$$-\frac{\eta_H+\eta_L}{4\sigma_0\sqrt{\Sigma}}\phi\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\sqrt{\Sigma}}\right)\left[1+\left(\frac{\bar{q}_{j0}-\bar{q}_{i0}}{2\sigma_0\sqrt{\Sigma}}\right)^2\right],$$

which is strictly negative. Thus, inequality (A.27) is satisfied for all values of $\bar{q}_{j0}-\bar{q}_{i0}$. □

Appendix B. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.geb.2021.11.007.

References

Allcott, H., Gentzkow, M., 2017. Social media and fake news in the 2016 election. J. Econ. Perspect. 31, 211–236.
Anbarci, N., Ghosh, S.P., Roy, J., 2017. Information control in reputational cheap talk. Games Econ. Behav. 106, 153–160.
Andina-Díaz, A., García-Martínez, J.A., 2021. Garbling an evaluation to retain an advantage. Discussion Paper.
Banerjee, S., 2021. Does competition improve analysts' forecast informativeness? Manag. Sci. 67, 3219–3238.
Baricz, Á., 2008. Mills' ratio: monotonicity patterns and functional inequalities. J. Math. Anal. Appl. 340, 1362–1370.
Beck, L., Witte, G., 2017. How Dirty Campaigning and Fake Facebook Sites Came to Dominate the Austrian Election. Washington Post, October 15.
Callander, S., 2008. Political motivations. Rev. Econ. Stud. 75, 671–697.
Crawford, V., Sobel, J., 1982. Strategic information transmission. Econometrica 50, 1431–1451.
DeGroot, M., 1970. Optimal Statistical Decisions. McGraw-Hill, New York.
Dellarocas, C., 2006. Strategic manipulation of Internet opinion forums: implications for consumers and firms. Manag. Sci. 52, 1577–1593.
Dubey, P., Haimanko, O., 2003. Optimal scrutiny in multi-period promotion tournaments. Games Econ. Behav. 42, 1–24.
Dubey, P., Wu, C.-w., 2001. Competitive prizes: when less scrutiny induces more effort. J. Math. Econ. 36, 311–336.
Duffy, M., Ganster, D., Pagon, M., 2002. Social undermining in the workplace. Acad. Manag. J. 45, 331–351.
Edmond, C., 2013. Information manipulation, coordination, and regime change. Rev. Econ. Stud. 80, 1422–1458.
Ely, J.C., Välimäki, J., 2003. Bad reputation. Q. J. Econ. 118, 785–814.
Fishburn, P., Gehrlein, W., 1977. Towards a theory of elections with probabilistic preferences. Econometrica 45, 1907–1924.
Gentzkow, M., Kamenica, E., 2014. Costly persuasion. Am. Econ. Rev. 104, 457–462.
Gentzkow, M., Kamenica, E., 2016. Competition in persuasion. Rev. Econ. Stud. 84, 300–322.
Gentzkow, M., Kamenica, E., 2017. Bayesian persuasion with multiple senders and rich signal spaces. Games Econ. Behav. 104, 411–429.
Grossman, G., Helpman, E., 1996. Electoral competition and special interest politics. Rev. Econ. Stud. 63, 265–286.
Grunewald, A., Hansen, E., Pönitzsch, G., 2020a. Political selection and the optimal concentration of political power. Econ. Theory 70, 273–311.
Grunewald, A., Klümper, A., Kräkel, M., 2020b. Welfare effects of information manipulation in political elections. working paper.
Grunewald, A., Kräkel, M., 2017. Advertising as signal jamming. Int. J. Ind. Organ. 55, 91–113.
Harmon, A., 2004. Amazon Glitch Unmasks War of Reviewers. New York Times, 14th of February.
Höffler, F., Sliwka, D., 2003. Do new brooms sweep clean? When and why dismissing a manager increases the subordinates' performance. Eur. Econ. Rev. 47, 877–890.
Holmström, B., 1999. Managerial incentive problems: a dynamic perspective. Rev. Econ. Stud. 66, 169–182.
Hsu, S.S., 2017. 'Pizzagate' Gunman Says He Was Foolish, Reckless, Mistaken – and Sorry. Washington Post, June 14.
Kamenica, E., 2019. Bayesian persuasion and information design. Annu. Rev. Econ. 11, 249–272.
Kamenica, E., Gentzkow, M., 2011. Bayesian persuasion. Am. Econ. Rev. 101, 2590–2615.
Kang, C., 2016. Fake News Onslaught Targets Pizzeria as Nest of Child-Trafficking. New York Times, November 16.
Lazear, E., Rosen, S., 1981. Rank-order tournaments as optimum labor contracts. J. Polit. Econ. 89 (5), 841–864.
Lazer, D.M.J., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S.A., Sunstein, C.R., Thorson, E.A., Watts, D.J., Zittrain, J.L., 2018. The science of fake news. Science 359 (6380), 1094–1096.
Levy, G., 2004. Anti-herding and strategic consultation. Eur. Econ. Rev. 48 (3), 503–525.
Little, A.T., 2012. Elections, fraud, and election monitoring in the shadow of revolution. Q. J. Polit. Sci. 7, 249–283.
Little, A.T., 2017. Propaganda and credulity. Games Econ. Behav. 102, 224–232.
Majumdar, S., Mukand, S., 2004. Policy gambles. Am. Econ. Rev. 94, 1207–1222.
Mayzlin, D., Dover, Y., Chevalier, J., 2014. Promotional reviews: an empirical investigation of online review manipulation. Am. Econ. Rev. 104, 2421–2455.
Meyer, M., Vickers, J., 1997. Performance comparisons and dynamic incentives. J. Polit. Econ. 105, 547–581.
Miklós-Thal, J., Ullrich, H., 2015. Belief precision and effort incentives in promotion contests. Econ. J. 125, 1952–1963.
Milgrom, P., 1988. Employment contracts, influence activities, and efficient organization design. J. Polit. Econ. 96, 42–60.
Milgrom, P., Roberts, J., 1988. An economic approach to influence activities in organizations. Am. J. Sociol. 94, S154–S179.
Morris, S., 2001. Political correctness. J. Polit. Econ. 109, 231–265.
Murphy, K.J., 1992. Performance measurement and appraisal: motivating managers to identify and reward performance. In: Bruns, W.J. Jr. (Ed.), Performance Measurement, Evaluation, and Incentives. Boston, pp. 37–62.
Nalebuff, B.J., Stiglitz, J.E., 1983. Prizes and incentives: towards a general theory of compensation and competition. Bell J. Econ. 14, 21–43.
Oltermann, P., 2017. Negative Campaign Sites Scandal Shakes up Austrian Election Race. The Guardian, October 5.
Ottaviani, M., Sørensen, P.N., 2006a. Reputational cheap talk. Rand J. Econ. 37, 155–175.
Ottaviani, M., Sørensen, P.N., 2006b. The strategy of professional forecasting. J. Financ. Econ. 81, 441–466.
Prat, A., 2005. The wrong kind of transparency. Am. Econ. Rev. 95, 862–877.
Ross, S., 2010. Introduction to Probability Models, 10th edn. Elsevier, Amsterdam.
Sanger, D.E., 2017. Putin Ordered 'Influence Campaign' Aimed at U.S. Election, Report Says. New York Times, January 6.
Shaked, A., Sutton, J., 1982. Relaxing price competition through product differentiation. Rev. Econ. Stud. 49, 3–13.
Shane, S., 2017. The Fake Americans Russia Created to Influence the Election. New York Times, September 7.
Stein, J., 1989. Efficient capital markets, inefficient firms: a model of myopic corporate behavior. Q. J. Econ. 104, 655–669.
Zinman, J., Zitzewitz, E., 2016. Wintertime for deceptive advertising? Am. Econ. J. Appl. Econ. 8, 177–192.

References (online appendix)

Baye, M.R., Hoppe, H.C., 2003. The strategic equivalence of rent-seeking, innovation, and patent-race games. Games Econ. Behav. 44, 217–226.
Baye, M.R., Kovenock, D., De Vries, C.G., 1996. The all-pay auction with complete information. Econ. Theory 8, 291–305.
Hillman, A.L., Riley, J.G., 1989. Politically contestable rents and transfers. Econ. Polit. 1, 17–39.
Konrad, K.A., 2009. Strategy and Dynamics in Contests. Oxford University Press, Oxford, New York.
Loury, G.C., 1979. Market structure and innovation. Q. J. Econ. 93, 395–410.
Tirole, J., 1988. The Theory of Industrial Organization. MIT Press, Cambridge, MA.
Tullock, G., 1980. Efficient rent seeking. In: Toward a Theory of the Rent-Seeking Society. Texas A & M Press, College Station, pp. 97–112.