The Leadership Quarterly xxx (xxxx) xxxx

Contents lists available at ScienceDirect

The Leadership Quarterly

journal homepage: www.elsevier.com/locate/leaqua

Full Length Article

How to prevent leadership hubris? Comparing competitive selections, lotteries, and their combination

Joël Berger[a,b], Margit Osterloh[c,d,e], Katja Rost[a,*], Thomas Ehrmann[f]

[a] University of Zurich, Department of Sociology, Andreasstr. 15 (5. OG), CH-8050 Zürich, Switzerland
[b] University of Bern, Department of Sociology, Fabrikstrasse 8, CH-3012 Bern, Switzerland
[c] University of Zürich, Switzerland
[d] University of Basel, Switzerland
[e] CREMA – Center for Research in Economics, Management and the Arts, Südstrasse 11, CH-8008 Zürich, Switzerland
[f] University of Münster, Institute for Strategic Management, Leonardo Campus 18, DE-48149 Münster, Germany

ARTICLE INFO

Keywords:
Leadership hubris
Appointment methods
Competitive selection
Lottery
Selection combined with lottery
Laboratory experiment

ABSTRACT

Hubris is a tendency of leaders to hold an overly confident view of their own capabilities and to abuse power for their own selfish goals, sometimes with disastrous consequences for organizations. A major reason for hubris is the rigorous selection process leaders typically undergo. This study proposes a governance mechanism used successfully in history to tackle hubris: partly random selections, which combine competitive selections by competence with lotteries. A frequently voiced concern about the use of lotteries is that it takes no account of the competence of the leader chosen. We propose that partly random selections can mitigate the disadvantages of both competitive selections alone and lotteries alone and reduce hubris in leaders. We conduct a test of this governance mechanism by means of a computerized laboratory experiment. Our results show that partly random selections significantly reduce the hubris of group leaders.

Introduction

As Aristotle famously noted, power and glorification often lead to hubris in leaders, which is defined as overconfidence in one's own abilities and the abuse of power (Aristotle Aristotle, 2003: 1378b 23–30). Hubris results in neglecting the limitations and precariousness of one's human condition (e.g. Cairns, 1996). For example, CEOs affected by hubris pay high premiums for unprofitable corporate acquisitions (Billett & Qian, 2008; Hayward & Hambrick, 1997a; Malmendier & Tate, 2008), invest in pet projects funded by internal cash flows (Malmendier & Tate, 2005), compensate themselves with salaries that the firm's performance does not justify (Billett & Qian, 2008), and demand rewards based on luck or other factors beyond their control (Liu & De Rond, 2016). Another example is the tax evasion revealed in the so-called Panama Papers, which were leaked in 2015. It became apparent that 143 politicians from all over the world, including many former and current democratic countries' heads of state and government were corrupted by their power and used offshore shell corporations to hide illegal financial transactions (Obermayer & Obermaier, 2016).

Previous research has extensively analyzed the detrimental consequences of hubris, in particular the consequences of CEO hubris for corporate outcomes. Governance mechanisms that may limit hubris of leaders have often been discussed, but mostly under the perspective of board control. For example, Hayward and Hambrick (1997a) show that vigilant boards, characterized by an independent board chair, outsider directors, and share ownership, restrain CEOs from paying large premiums for corporate acquisitions, which they might otherwise feel hubristically entitled to do.

To our knowledge, this is the first study that proposes a governance mechanism used successfully in history to tackle the problem of leadership hubris today: competitive selection combined with lotteries (in the following, termed *partly random selection*). Historical evidence suggests that being chosen randomly prevents hubris, which is overconfidence and the abuse of power (Buchstein, 2010; Dowlen, 2017; Duxbury, 2002; Manin, 1997; McCormick, 2006; Sintomer, 2014; Van Reybrouck, 2016). It has also been shown that successful people who recognize that randomness or luck has played an important role more often express humility and a pro-social focus (Bartlett & DeSteno, 2006; Frank, 2016). However, lotteries may result in incompetent candidates being chosen. Therefore, they have usually been combined with conventional selection methods. In this vein, we develop our suggestion.

* Corresponding author.
E-mail address: katja.rost@uzh.ch (K. Rost).

https://doi.org/10.1016/j.leaqua.2020.101388
Received 18 April 2018; Received in revised form 23 September 2019; Accepted 24 September 2019
1048-9843/

Including randomness in the process through which leaders are appointed raises awareness among leaders themselves, board members, journalists, and others that "great man" theories may be unrealistic.

Lotteries today are rarely used to appoint leaders in real-life organizations, in contrast to what we know from history (see e.g. Buchstein, 2009; Manin, 1997; Sintomer, 2014; Van Reybrouck, 2016). Therefore, to test our theory, we devised a laboratory experiment in which subjects are appointed as group leaders in three treatment conditions: a competitive selection treatment, a random selection treatment, and partly random selection, in which competitive selection is combined with lottery. The design of this experiment is inspired by a historical example at the University of Basel, called *Wahl zu Dreyen* or selection from three (Burckhardt, 1916: 34; Stolz, 1986: 670). We measure hubris by using an incentivized behavioral measure, which has been applied in former experimental research on the abuse of power (Bendahan, Zehnder, Pralong, & Antonakis, 2015) and which has been suggested in observational studies to be a good proxy for leadership hubris (Hayward & Hambrick, 1997a). This design allows us to test whether the historically approved partly random selection of leaders helps to reduce leadership hubris more than the current practice of selecting leaders on the basis of competitive performance evaluations.

In contrast to former studies on leadership hubris, our study focusses on how features of organizational design shape the incentives of leaders (Zehnder, Herz, & Bonardi, 2017). The strength of this economic approach is that it determines the exogeneous conditions under which leaders with hubris can and will emerge. Former studies analyze the psychological traits of leaders or observe the performance outcomes of leaders; the empirical results are thus confounded by selection effects and omitted variables and are prone to endogeneity bias. Former research on leadership hubris does also not view organizational choices as trade-offs-decisions requiring a cost-benefit-analysis in which one option is compared to another. In our theory and experimental design, we consider that the cost and benefits of alternative methods of leader selection depend on the situation at hand. It implies that there is no globally best alternative of leader selection; but a best alternative to prevent leadership hubris. To study the exogenous conditions of the emergence of leadership hubris and to compare the different alternative selection methods against each other, we set up a microeconomic structure, where individuals are randomized to treatment, with clear and consequential links between decisions and payouts (Zizzo, 2010).

The remainder of the paper is structured into four sections. In the next section, we define leadership hubris and discuss how a random component in competitive selections can help to reduce leadership hubris. We present our main arguments with a formal model comparing competitive selection methods with partly random selection methods. We model the tradeoffs between the positive competence effects of competitive selections and the positive effects of partly random selections in preventing power abuse. Then, we introduce the empirical strategy we applied to test our hypothesis in a laboratory experiment and present our empirical findings. We close our paper with a discussion of our findings in the context of the existing literature on leadership hubris and their implications for the practice of selecting leaders.

Selection methods as a trigger for leadership hubris

What is leadership hubris?

The concept of hubris has attracted growing interest in business and management research, yet it is poorly defined in this literature (Hollow, 2014). Richard Roll (Roll, 1986) was the first to use the term to show that hubristic CEOs are often responsible for shareholder losses during mergers and acquisitions. Subsequently, researchers in management and finance have used the hubris concept to demonstrate the negative consequences of "exaggerated self-confidence" in CEOs (Chatterjee & Hambrick, 2007; Hayward & Hambrick, 1997a; Hayward, Shepherd, & Griffin, 2006; Hiller & Hambrick, 2005; Malmendier & Taylor, 2015; Roll, 1986). For example, it has been shown that hubristic CEOs are more likely to destroy their firm's value (Bertrand & Mullainathan, 2001; Billett & Qian, 2008; Hayward & Hambrick, 1997a; Malmendier & Tate, 2005, 2008), violate integrity standards (Blickle, Schlegel, Fassbender, & Klein, 2006; O'Connor, Mumford, Clifton, Gessner, & Connelly, 1996), have a negative impact on employee satisfaction (Blair, Hoffman, & Helland, 2008), and overestimate the expected returns from investing in new products and markets (Camerer & Lovallo, 1999). Whereas the studies convincingly demonstrate the undesirable effects of hubris, they fail to provide a definition of it (Petit & Bollaert, 2012).

Some studies proxy hubris as mental concepts such as overconfidence or narcissism. Overconfidence is defined as individuals' overestimation of their own abilities and the outcomes of these when assessing their skills (Larwood & Whittaker, 1977; Moore & Healy, 2008).[1] Narcissism is a mental condition in which people have an inflated sense of their own importance, a strong need for excessive attention and admiration, troubled relationships, and a lack of empathy for others (Caligor, Levy, & Yeomans, 2015). It is not clear in these studies whether, and if so how, hubris differs from overconfidence or narcissism, whether it is essentially a mental concept of motivated cognition (Bénabou & Tirole, 2016) or also includes behavioral characteristics, and how it develops over time. Because of this lack of clarity, the management and finance literature provides no satisfactory scope for understanding when and how preventive action might be taken (Petit & Bollaert, 2012).

To overcome these issues, Petit and Bollaert (2012) brought together work on hubris from the fields of mythology, psychology, and ethics to develop a framework for defining leadership hubris. In Greek tragedies, hubris described arrogant or foolish behavior that belies common sense and results from excessive pride and self-confidence and a loss of contact with reality (Nicolas & Fisher, 1992). The Ancient Greek view is that hubris "is not only an attitude, it is a kind of action as well" (Woodruff, 2005: 15). In psychology, the term describes how certain leaders, democratically elected heads of government as well as dictators, behave with irrational self-confidence in their own abilities when put in positions of immense power, with harmful results for the wider population (Owen, 2012; Owen & Davidson, 2009). According to this literature, hubris is a latent personality trait that is triggered by accession to a position of power and the resulting lack of constraints on the person's behavior (Petit & Bollaert, 2012). Criteria for diagnosing hubris syndrome include a narcissistic propensity to see the world primarily as an arena in which to exercise power and seek glory, overconfidence in his or her own judgment, contempt for the advice and criticism of others, a disproportionate concern with image and presentation, an identification with the organization or country to the extent that the person regards his or her interest as identical with that community, (Owen, 2012; Owen & Davidson, 2009). This list shows that the hubris syndrome in psychology overlaps considerably with overconfidence and narcissism and is a combination of both attitudes and behaviors. Whereas overconfidence and narcissism are relatively stable personality traits, hubris is more dynamic and most importantly is activated by an environmental trigger, in particular accession to a position of power (Winnington-Ingram, 1971).

Petit and Bollaert (2012) conclude from their analysis that leadership hubris is present solely in a context of power and has both cognitive and behavioral aspects. Cognitive aspects of hubris include the

[1] In the management literature, overconfidence is mostly defined according to its measured outcomes rather than its psychological roots. Measures are, for instance, the degree to which CEOs do not diversify their risks with respect to firm-specific assets because they overestimate the performance of their firm (see e.g. Malmendier & Taylor, 2015). Consequently, no differentiation is made between overestimation of one's ability, belief in being better than average, or certainty about knowing correct answers, see Moore & Healy, 2008.

overestimation of one's own abilities, and likelihood of success. Behavioral aspects of hubris center on corruption of power, defined as the extent to which leaders use their power for their personal gain and/or contravene social norms to benefit themselves to the detriment of social welfare (Bendahan et al., 2015). In line with this literature, we define leadership hubris as follows:

> *Hubris is the abuse of power by individuals who are overconfident and, on gaining positions of power, benefit themselves to the detriment of other members of the community.*

This definition of hubris considers both cognitive and behavioral aspects in the context of power. Power is defined as “having the discretion and the means to asymmetrical enforce one's will over others” (Sturm & Antoniakis, 2014: 139).[2] The cognitive and behavioral parts of the definition are clearly differentiated and can easily be measured in laboratory experiment (and, with regard to future studies, also with observational field studies).

In the following, we discuss how selection methods may trigger hubris in overconfident persons promoted to leadership positions. In contrast to Bendahan et al. (2015), who study the impact of *given* levels of power on corrupt behavior under various conditions, we investigate how hubris *emerges* in the context of power following different selection methods, competitive and partly random selection, and its impact on corrupt behavior.

Why competitive selection methods may foster hubris

Competitive selection methods are applied to select the most competent person, usually the person with the best performance record as a leader. Such methods prevail in business organizations, governments, and other organizations. We argue that this method triggers leadership hubris of overconfident people when they are selected as leaders due to two factors.

First, some persons have a predisposition for hubris because they show higher self-serving bias (Dobbins & Russell, 1986; Miller & Ross, 1975) respectively fundamental attribution error (Barber & Odean, 2001; Bénabou & Tirole, 2016; Camerer & Lovallo, 1999; Hayward & Hambrick, 1997a; Hayward et al., 2006; Hiller & Hambrick, 2005; Roll, 1986). It suggests that people tend to overattribute certain behaviors and outcomes to their own disposition and underestimate the influence of situational forces and luck (Gilbert & Malone, 1995; Liu & De Rond, 2016; Weber, Camerer, Rottenstreich, & Knez, 2001). It has been for example shown that positions of power are often occupied by persons with higher self-serving bias respectively fundamental attribution error (Brunell et al., 2008; Nevicka, Ten Velden, De Hoogh, & Van Vianen, 2011; Paunonen, Lönnqvist, Verkasalo, Leikas, & Nissinen, 2006). Second, competitive selection methods confirm overconfident leaders' feeling that they are exceptional and perform far “above the average” of other candidates (Alicke, 1985; Brown, 1986; Malmendier & Taylor, 2015; Miller & Ross, 1975). They feel superior to their subordinates and others and see those people as means to satisfying their personal ends (Keltner, Gruenfeld, & Anderson, 2003). In short, competitive selections can trigger latent tendencies for hubris in people with a respective disposition.

Random selection as an alternative selection method in history

History shows that alternative selection methods have been extensively used to appoint leaders. Lotteries, in the following called *random selection methods*, have played a considerable role in history, a fact that has recently gained some attention in political science (Buchstein, 2009, 2010; Dowlen, 2017; Duxbury, 2002; Frey & Osterloh, 2016; Manin, 1997; McCormick, 2006; Sintomer, 2014; Van Reybrouck, 2016). In classical Athens, random selection was the method by which the members of the *boulê*,[3] judges, and magistrates were appointed (Duxbury, 2002; Hansen, 1991).

The most important advantage of random selection methods is that it prevents aristocratic effects: it prevents individuals and influential family dynasties from monopolizing power. Wealthy and powerful people have better chances of standing for office and of accumulating resources that help them to dominate the political process and to win elections (McCormick, 2006). In addition, random selections lead to humility[4] and help to prevent internal political conflicts.[5] The main disadvantage of random selection, termed amateurism, is that candidates chosen by lottery may turn out to be incompetent to perform the tasks in question (Manin, 1997; Van Reybrouck, 2016).

To mitigate this disadvantage, in the late Middle Ages random selection and competitive selection methods were combined in what are here called *partly random selection methods* (see e.g. Frey & Osterloh, 2016; Zeitoun, Osterloh, & Frey, 2014). After the expulsion of the Medici from Florence in 1494, the city introduced partly random selection methods around 1497–1499 to appoint the magistrate (McCormick, 2006; Pocock, 2016; Sintomer, 2014).[6] In Venice, the procedure by which the Doge—the chief magistrate and leader of the Republic of Venice—was appointed began with a lottery that included all of the members of the Grand Council to determine an electoral body. The initial lottery was followed by several rounds in which competitive selection and random selection alternated. Finally, the Doge was appointed by competitive selection. Parma, Bologna, Bern, Frankfurt, and many other communities also used a combination of random and competitive selection to appoint their executives (Buchstein, 2009).

Historically, the trade-off between amateurism and aristocratic effects has been particularly relevant in expert organizations. Office holders are selected because of their professional knowledge. Nevertheless, aristocratic criteria may play an unwarranted role if influential experts deploy their power to influence the appointment of like-minded individuals or otherwise exploit their professional networks. This problem led to changes that took place at the University of Basel in the 18th century. Until the end of the 17th century, the appointment of professors at this university was seriously compromised by the interventions of politically influential family dynasties and by corruption. To combat this problem, a law was passed requiring the appointment of new professors through a procedure that combined competitive and random selection (Burckhardt, 1916), termed the *Wahl zu Dreyen* or selection from three. The law, which was introduced in 1718, required candidates to submit proof of their qualifications to the governing body of the university, which then decided whether a candidate was eligible or not (Burckhardt, 1916). Subsequently, all the

[2] Because we focus on the behavioral effects of power, we do not consider the psychological processes of how power changes the powerholder, for example by power disinhibition or power approach (see e.g. Keltner et al., 2003; Lammers, Galinsky, Dubois, & Rucker, 2015).

[3] The *boulê* was a council of 500 citizens who prepared the general assembly of the people.

[4] The generation of humility with officials in ancient Greece is dealt with in Duxbury, 1999: 28–29.

[5] A historical example showing that partly random selection has successfully contributed to avoiding detrimental political conflicts is demonstrated by Greif (1995). He compares the political organizations of Genoa and Venice during the medieval ages. Genoa, which introduced partly random selection much later than Venice, declined compared to Venice. Genoa suffered from many internal civil wars between the different powerful clans, whereas Venice enjoyed high political stability for centuries.

[6] This constitutional innovation expanded the pool of possible office-holders from the 200 members of the wealthiest families, as had been the case up to that point, to approximately 3200 citizens. Candidates (*nominati*) were proposed by the noble families and guilds. Those who were considered to be suitable for office were selected from among the proposed candidates. The procedure ended with a lottery (*la tratta*) (see e.g. Van Reybrouck, 2016).

professors of the University of Basel came together to act as the electoral authority. If two or three candidates were eligible, the candidate to be appointed was chosen by lottery. If more than three candidates came into question, the electoral authority was divided by lottery into three electoral colleges. Each college had to propose one candidate by secret voting. Finally, the candidate to be appointed was decided by lottery.

Today, few organizations apply random and partly random selection methods to appoint leaders. The Coptic Pope is appointed by random selection from three candidates, the Amish choose their leaders randomly, and random selection is commonly used as a mechanism to select juries or decide tiebreaks in national and local elections (Boochs, 2004; Frey & Steiner, 2014; Shoucri, 1991). University places or grants are sometimes partially allocated by random selection methods.[7] It has been suggested to apply partly random selection to the selection of papers to be published (Osterloh and Frey, in press; Oswald, in press). In Switzerland recently, a popular initiative has been started to draw federal judges from a pool selected by an expert committee.[8] However, modern organizations rarely use partly random selection methods.

How partly random selection methods can help to avoid hubris

We have argued that overconfident individuals elected in competitive selections show a high tendency to misuse power to the detriment of other people or the common good, because such selections first reinforce self-serving bias respectively fundamental attribution errors and second may trigger hubris.

In contrast, selection methods with a random component do not send such signals because they appoint leaders by mathematical laws.[9] It is clear that luck, rather than individual performance, plays a crucial role in deciding who finally becomes leader. Self-serving bias and fundamental attribution error, in which individuals overattribute outcomes to their own performance, are impeded. Overconfident individuals are not confirmed in their belief that they are far above the average. Hubris is not activated.

> *Hypothesis: Compared to partly random selection, competitive selections foster hubris in leaders so that overconfident leaders will show a higher tendency to abuse their power to benefit themselves to the detriment of other people or the common good.*

Formal account of competitive selections as a trigger for leadership hubris

To better understand the mechanism through which hubris occurs, we provide a more formal, descriptive account of the economic effects of selection methods on the common welfare of the group in question (see the model with propositions and proofs in the Appendix 2). The analysis focuses on the tradeoffs between the gross positive competence effects of competitive selections and the positive effects of preventing the misuse of power by partly random selections. We model these determinants to establish a basic framework for the experimental work.

In general, we assume that groups screen leaders before hiring them to secure common welfare, for example the welfare of an organization, firm, or state. We define welfare as the difference between advantages of high leadership productivity gained by rigorous selection minus the costs of misuse of power, for example due to hubris, and an average market wage W. To maximize its welfare a firm has to solve a two-step selection problem using the screen: first it has to find the *highly productive* managers, then in the second step out of this pool the *good behavior* managers with no misuse of power need to be selected. For simplicity, we assume the pool of leaders in the population consists of types with either *high* or *low competence* with a particular distribution in the population. They also exhibit dispositions for either *good* or *bad behavior*, leading eventually to the abuse of power. The selection task gets complicated by information problems about productivity and managerial behavior the firm faces.

We hypothesize that a leader with strong overconfidence who went through a competitive selection process will act corruptly and vice versa. The trigger *(1 − t)* works depending on the distribution of a psychic disposition in the population of leaders. The selection process can be viewed as a trigger for this sort of behavior, which either hides or brings to surface the underlying dispositions. To model the consequences of this hypothesis, we compare the common welfare of the group in three selection processes: purely random (with zero screening costs), which is used as a *point of reference*, competitive selection, and partly random selection.

Common welfare of purely random selection versus competitive selection

The lack of a performance screen in a pure random selection has unavoidable negative productivity effects that in all relevant cases is not overcompensated by positive (anti) "hubris" effects. We assume a zero profit condition for pure random selection. Comparing common welfare of the competitive selection with pure random selection, only *the quality of the screen* can be chosen by the firm to improve its common welfare. Simple simulations show that even a high ratio of *"bad behavior"* managers of 60% (80%) is compensated profit-wise by a minimum screen quality of 58% (73%). We conjecture and proof (see proposition 1) for fairly general assumptions that when screening costs are sufficiently small and the quality of the screen is sufficiently high, competitive selection *always* dominates pure random selection.

Common welfare of competitive selection versus partly random selection

Partly random selection implies that *low competence* leaders are being selected with a higher probability than in the competitive selection case. But this loss of productivity will be completely overcompensated by the elimination of the trigger for bad behavior. Comparing common welfare of the competitive selection with partly random selection and holding the quality of the screen constant, implies that the change of selection probability *(r^*)* can be deliberately chosen by the firm to improve its common welfare. We conjecture and proof (see proposition 2) that an appropriately designed partly random selection always dominates pure competitive selection as a screening strategy. This result is similar to the selective intervention idea by Oliver Williamson ((1985): 133–138). Basically, the smaller the *(1 − t)* in the population the smaller r^* has to be that improves welfare. For small *(1 − t)* this change has to become very small implying that the selection process of the firm basically approaches the competitive selection process in reality. This result again can be illustrated with a numerical example from simple simulations. Given a ratio of 30% selfish and hyper selfish managers in the population – which is the result of the experiments - and assuming quality of the screen of 60% an extra of random selection of $r^* < 0{,}35$ leads to a higher profit in the partly random case than in the competitive case. Increasing this "*bad behavior"* ratio to 60% necessitates an $r^* < 0{,}575$ to compensate the hubris effects.

[7] Examples are the Health Research Council of New Zealand (2017) and the German Volkswagen Foundation, See https://www.volkswagenstiftung.de/experiment.htm.

[8] See https://www.bk.admin.ch/ch/d/pore/vi/vis486.html, retrieved July 17, 2019.

[9] In former times, lots were very often considered as a means of getting God to speak, most prominently in the Old Testament (for a discussion see Buchstein, 2019). However, as early as in ancient Athens, random selection was used not only for religious reasons, but because Athenians associated it with rationality and political virtues, in particular the affirmation of faith in democracy, political humility of officials, and the capacity of lots to mitigate conflicts (see Duxbury, 1999: 16–23).

An experimental test of the effect of alternative selection methods on leadership hubris

Experimental method

We test our theory in a laboratory experiment. In practice, conducting such an empirical test is not possible, because today modern organizations do not use lotteries to appoint their leaders. Our study complements the historical case studies, which derive from real-world settings but are not causally interpretable.

Subjects

We conducted a computerized laboratory experiment at the Decision Science Laboratory at ETH Zurich using zTree (Fischbacher, 2007). The participants, 864 students of the University of Zurich and ETH Zurich, were randomly selected from a pool of students who had volunteered to participate in behavioral experiments for monetary compensation. Participants on average gained USD 30 for 45 min. Some 61% of the sample were females. The mean age of participants was 23 years with a standard deviation of 3.7 years.

Treatment conditions

The 864 participants were randomly selected into groups of six and randomly assigned to one of three treatment conditions (between-subjects design with 48 groups per treatment condition). Out of the six group members one individual was selected as group leader through a treatment-specific selection method. The participants only learned about the selection method in their own treatment condition. In the *competitive treatment* condition, in each group the participant with the highest test score in a competence task was appointed as a group leader. The task involved giving correct answers to 30 general-information questions that covered a wide variety of topics, including history, sports, art, geography, literature, and entertainment (for example, “please name the capital of Vietnam”) under time pressure (Nelson & Narens, 1980); the questions had been used in previous laboratory experiments on methods of employee selection (Larkin & Leider, 2012). In the *random treatment* condition, one of the six group members was randomly selected as a group leader. In the *partly random treatment* condition, the three highest performing individuals were preselected, and then the group leader was randomly selected from the three preselected individuals. This selection mechanism mimics the historical example from the University of Basel (*Wahl zu Dreyen*), as documented by Burckhardt (1916).

Measures

Our definition of hubris considers cognitive and behavioral aspects, specifically overestimation of one's own abilities and abuse of power by benefiting oneself to the detriment of other members of the community. Consequently, we use two key dependent measures to operationalize hubris. The cognitive aspect of hubris is operationalized similar to previous experimental studies using the concept of overconfidence (Larkin & Leider, 2012) as the difference between participants' estimated number of correctly answered questions minus the number of objectively correctly answered questions. The behavioral aspect of hubris is measured with a variant of the dictator game (Kahneman, Knetsch, & Thaler, 1986) as applied by Bendahan et al. (2015). In this game, the “dictator”—called “group leader” in our experiment—decides how to split a pot of money between themselves and five subordinates. The subordinates have no say in this split. The group leader has four options. Option 2 was labeled as “default option” in order to provide an anchor point, but not a social norm (Bendahan et al., 2015). In this option, the leader received 220 monetary units (MP; one MP unit is approximately 0.11 USD), and each of the five followers received 190 MP. The remaining four options were labeled as “Option 1,” “Option 3,” and “Option 4”. Option 1 (prosocial option) provided an equal payoff of 210 MP to all six group members. In Option 3 (anti-social option), leaders allocated 270 to themselves, leaving 130 to each of their five subordinates. In Option 4 (very anti-social option) leaders misused their power greatly. They kept 370 MP, leaving only 10 MP to each of their subordinates. Importantly, the social behavior of the leader decreased from Option 1 to Option 4. The total sum was explicitly mentioned for all of the four options.

Whereas we expect overconfidence to play a crucial role in generating hubris, other factors could impact leaders abuse of power as well. First, the three selection methods could also engender different *distribution norms* – beliefs about what the proper behavior under a given selection regime is. For example, a competitive selection method based on competence could increase the group leader's legitimacy, and therefore, they could feel entitled to claim a bigger slice of the pie. For this reason, we endogenously measured the distribution norm in each group. Before the participants knew who would be appointed as a group leader, they answered the question how the group leader should distribute the money in the dictator game. We presented four different distribution norms: Option 1 (220 MP for the leader, 190 MP for each follower), Option 2 (210 MP for the leader and for each follower), Option 3 (270 MP for the leader, 130 MP for each follower), and Option 4 (370 MP for the leader, 10 MP for each follower). The most frequently chosen option in each group was defined as the valid distribution norm for that group. As in Bendahan et al.'s (2015) experiment, the distribution norm in the overwhelming majority of all groups was Option 2 (90.3%). Only 6.3% decided for Option 1 and 3.5% for Option 3. No group decided for Option 4.

Selfishness might be a further key factor that impacts the group leaders' distributional decisions, with more selfish group leaders allocating larger amounts to themselves. Similarly to Bendahan et al. (2015), we used the prisoner's dilemma game described in Fig. 1 as a behavioral measure of selfishness.

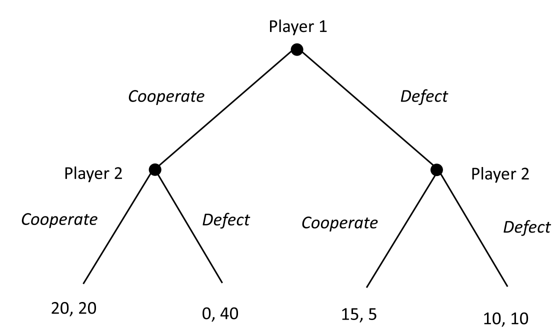

Fig. 1. Parameters of the prisoner's dilemma game (i.e. behavioral selfishness measure). As a behavioral measure of selfishness, participants play a one-shot, sequential prisoner's dilemma game with real monetary payoffs. We elicited second mover response using the strategy method (i.e. participants made choices in both the role of Player 1 and Player 2 before knowing to which role they ultimately would be randomly assigned). Player 1 decided whether or not to transfer their endowment of 10 monetary units (MU, with 1 MU = 1/9 USD approximately) to Player 2. The transferred amount would then be tripled by the experimenters and sent to Player 2. Player 2 then decided whether or not to send half of their earnings (i.e. their initial endowment plus the tripled transfer) back to Player 1. We used Player 2's cooperative response to Player 1's cooperation as a measure of selfishness. We do not use Player 1's decision. Because this decision reflects his or her trust in Player 2. The selfish decision for Player 2 is to not cooperate whereas to cooperate is the pro-social decision. The payoffs are determined as follows. Suppose Player 1 cooperates by transferring their initial 10 MU; these points are tripled thus giving Player 2 a total of 40 MU (30 MU from Player 1 plus Player 2's initial endowment of 10 MU). If Player 2 does not cooperate, they keep these 40 MU and Player 1 gets nothing. If Player 2 cooperates they can send half of their total earning back to Player 1 (i.e., 40/2 = 20 MU), giving both Player 1 and Player 2 each 20 MU. If both players defect, they both keep their initial endowment of 10 MU. If Player 1 defects (sends nothing) and Player 2 cooperates by sending half their endowment (i.e., 10/2 = 5 MU) to Player 1, the latter has 15 MU and Player 2 only 5 MU.

As control variables, we included perceived competitiveness of the selection method,[10] perceived significance of competence for being selected as a leader,[11] relative task performance in the competence test, risk seeking (scale from 0 to 10),[12] and gender.

Descriptive statistics and bivariate correlations are referred in Table 1.

Experimental procedure

The experiment proceeded as follows. First, selfishness was measured using the prisoner dilemma's game involving a second mover's reaction to a first mover's cooperation (selfish or pro-social reaction), as described in Fig. 1. Second, the competence task was conducted. Third, the participants learned about the method of leader selection, but not yet about the outcome of the selection process. Fourth, we endogenously measured the group-specific distribution norm. Fifth, the participants were informed about the group-specific norm for proper leader behavior, which was the option most frequently chosen in each group. Sixth, the participants learned whether or not they had been appointed as group leader. Seventh, we measured subjective task performance to differentiate between overconfident and underconfident individuals (note: individuals did not know their objective task performance from the competence test). Further, we measured risk preferences, perceived competitiveness in the selection process, and perceived level of competence required for selection as a group leader. Eighth, the group leader decided how to split the sum. Ninth, in addition to the sum the leader distributed, participants received a show-up fee of 10 USD and the money received in prisoner dilemma's game, which was undertaken to test selfishness. The experiment ended with a brief questionnaire on the socio-demographic background of the participants.

Results

In the experiment, 144 of the 864 participants were selected as leaders, i.e. in each of the three treatments, we observe 48 leaders. Before we test the hypotheses, we analyze the competence differences between followers and leader in the different treatments. Table A1 in the appendix documents the pure descriptive results. For each treatment it shows the number of correctly solved tasks of all participants and of group leaders and the contrasts in competence scores between the treatments. To test how competence differs between leaders and followers in the treatments, Table 2 documents a regression analysis with competence as the dependent variable and treatment, group leader and the interaction between treatment and group leader as independent variables, controlling for the exogeneous and endogenous variables of the final model. Table 2 indicates no significant difference between the competence scores of participants in the treatments (Model 1), but that group leaders on average perform significantly better than followers with around 4.313 more correctly solved tasks (Model 2). The interaction effects show that competences between leaders and followers are significant different in the treatments (Model 3). Fig. 2 graphically illustrates the results. Compared to their followers, leaders in the competitive treatment solved almost 7 more tasks correctly. Leaders in the partly random treatment solved around 3.5 more tasks correctly than their followers. In the random treatment, the performance between leaders and followers does converge; leaders solved only 1.5 more tasks correctly. The observations are consistent with our theory and formal

[10] "How competitive did you perceive the group manager selection process?" 0 = not competitive at all up to 10 = very competitive.

[11] "How big is the influence of a person's competence on the chance to become a group manager?" 0 = no influence at all up to 10 = very strong influence.

[12] "How willing are you to take risks, in general?" 0 = not willing at all up to 10 = perfectly willing (Dohmen, Falk, Huffman, & Sunde, 2011).

Table 1
Correlation matrix.

ID	Variable	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Competitive treatment	0.333	0.472	1.000													
2	Partly random treatment	0.333	0.472	−0.500	1.000												
3	Random treatment	0.333	0.472	−0.500	−0.500	1.000											
4	Selfish decision in the dictator game (1–4)	0.292	0.456	0.032	−0.032	0.000	1.000										
5	Overconfidence	0.334	3.549	0.028	0.013	−0.041	0.013	1.000									
6	Overconfidence × Competitive treatment	0.159	2.102	0.107	−0.053	−0.053	0.180	0.588	1.000								
7	Overconfidence × Partly random treatment	0.133	1.999	−0.047	0.094	−0.047	−0.210	0.560	−0.005	1.000							
8	Overconfidence × Random treatment	0.043	2.062	−0.015	−0.015	0.029	0.035	0.579	−0.002	−0.001	1.000						
9	Prosocial norm	1.972	0.311	−0.215	0.083	0.133	−0.254	−0.012	−0.068	0.008	0.041	1.000					
10	Selfishness	0.456	0.498	−0.016	0.013	0.003	0.408	−0.025	−0.010	−0.037	0.004	0.026	1.000				
11	Risk seeking	3.010	1.014	−0.034	0.065	−0.032	0.230	0.080	0.068	0.004	0.064	−0.027	0.025	1.000			
12	Female	0.611	0.488	−0.030	0.035	−0.005	−0.149	−0.037	−0.045	−0.010	−0.008	0.019	−0.051	−0.090	1.000		
13	Competitive	6.082	2.879	0.210	0.054	−0.264	−0.036	0.072	0.030	0.022	0.073	−0.076	0.050	0.061	−0.003	1.000	
14	Solved Problems	9.545	4.597	0.010	−0.071	0.061	−0.081	−0.122	−0.054	−0.053	−0.104	−0.017	0.019	−0.132	−0.165	−0.013	1.000

Table 2
Regression models to predict scores in the competence test (measured by the number of solved tasks).

	Model 1	Model 2	Model 3	Model 4	Model 5
Competitive treatment	(Ref.)	(Ref.)	(Ref.)	(Ref.)	(Ref.)
Random treatment	0.337	0.337*	1.158**	1.172**	1.086**
	(0.88)	(0.94)	(3.00)	(3.04)	(2.74)
Partly random treatment	−0.524	−0.524	0.058	0.206	0.204
	(−1.37)	(−1.46)	(0.15)	(0.53)	(0.53)
Group leader		4.313***	7.120***	6.819***	6.876***
		(10.99)	(10.64)	(10.32)	(10.38)
Competitive treatment × Group leader			(Ref.)	(Ref.)	(Ref.)
Random treatment × Group leader			−4.929***	−4.601***	−4.725***
			(−5.21)	(−4.94)	(−5.05)
Partly random treatment × Group leader			−3.495***	−3.379***	−3.423***
			(−3.69)	(−3.63)	(−3.68)
Prosocial Groupnorm				−0.316	−0.359
				(−0.65)	(−0.73)
Pro-social preferences				(Ref.)	(Ref.)
Selfish Preferences				0.290	0.310
				(1.02)	(1.09)
Risk seeking				−0.469**	−0.449**
				(−3.31)	(−3.16)
Male				(Ref.)	(Ref.)
Female				−1.475***	−1.436***
				(−5.04)	(−4.90)
Perceived Competitiveness					0.024
					(0.43)
Perceived Competence					−0.086
					(−1.70)
Constant	9.608***	8.889***	8.421***	10.860***	11.206***
	(35.54)	(33.98)	(30.81)	(15.60)	(14.37)
Adj R-squared	0.0036	0.1253	0.1516	0.1815	0.1825
N	864	864	864	864	864

z statistics in parentheses.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

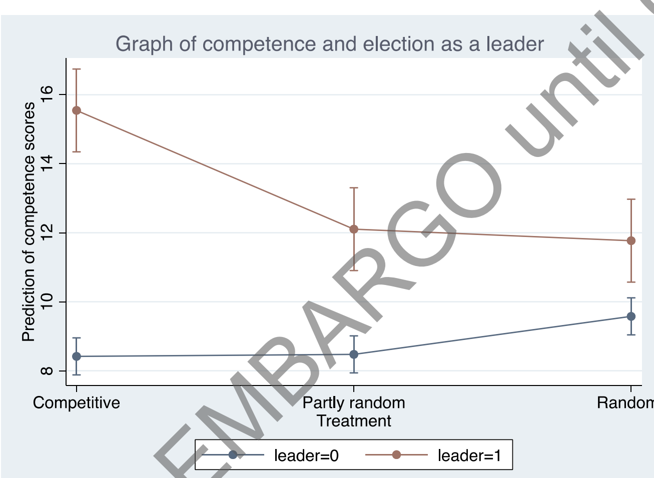

Fig. 2. Marginsplot of the interaction between treatment and election as a group leader respectively follower to predict scores in the competence test measured by the number of solved tasks in the different treatments (estimations of Table 2, Model 3). The graph illustrates that competence differences between group leaders and followers are highest in the competitive treatment managers, lowest the random treatment and medium in the partly random treatment.

model. It suggests that, compared to competitive selections, a random component in selections increases lower competence of leaders. However, compared to purely random selections, partial random selections lead to more competent leaders.

It should be mentioned that the randomization in our experiment worked good but not perfectly. First, as indicated in Table A1 and Fig. 2 leaders in the random treatment perform slightly better than their followers. Ideally in the random treatment the competence of leaders and followers should be equal. Second, Table A1 indicates that participants in the partial random treatment had slightly lower competence scores as the participants in the random treatment. Theoretically there should be no difference in competence scores between the participants in both treatments. We nevertheless belief that both small deviations from the ideal situation are unproblematic for the following analyses: it may lead to an underestimation but not an overestimation of our hypothesis. Compared to random selections, the competences of leaders in partial random selections are in reality even higher than in our experiment. The deviations have no impact on our core comparison between the competitive treatment and the partial random treatment.

To test whether overconfident group leaders are more prone to hubris in competitive selections than in partly random selections, we measure whether those leaders abuse their power to benefit themselves to the detriment of their followers. Table A2 in the appendix gives a pure descriptive overview of the findings. The table presents the selfishness of decisions of overconfident leaders, i.e. pro-social decisions (option 1), standard decisions (option 2), selfish decisions (option 3) and hyperselfish decisions (option 4). For descriptive reasons in Table A2 we defined overconfident leaders as persons that overestimated their correctly solved tasks, i.e. the difference between participants' estimated number of correctly answered questions minus the number of objectively correctly answered questions is greater than “0”. In the following regressions overconfidence will be included on a metric scale. Table A2 reveals that we observe the highest number of overconfident leaders (28/58%) in the competitive treatment, followed by the partial random treatment (25/52%) and the random treatment (17/35%). The small number of overconfident leaders in the random treatment can be explained by the higher percentage of low performers in this treatment compared to selections with a competitive element. Low performing

Table 3
Binary logit regression models to predict selfish decisions of leaders (1: hyperselfish/selfish decision, 0 standard/pro-social decision).

	Model 1	Model 2	Model 3	Model 4	Model 5
Competitive treatment	(Ref.)	(Ref.)	(Ref.)	(Ref.)	(Ref.)
Random treatment	−0.099	−0.088	0.224	0.439	−0.974
	(−0.22)	(−0.19)	(0.45)	(0.71)	(−1.18)
Partly random treatment	−0.202	−0.202	0.239	0.223	−0.346
	(−0.45)	(−0.45)	(0.46)	(0.35)	(−0.50)
Overconfidence		0.009	0.205*	0.168	0.150
		(0.16)	(2.00)	(1.28)	(1.16)
Competitive treatment × Overconfidence			(Ref.)	(Ref.)	(Ref.)
Random treatment × Overconfidence			−0.167	−0.178	−0.140
			(−1.22)	(−1.06)	(−0.78)
Partly random treatment × Overconfidence			−0.541**	−0.533*	−0.553*
			(−3.09)	(−2.38)	(−2.37)
Prosocial Groupnorm				−2.360***	−2.815***
				(−3.15)	(−3.46)
Pro-social preferences				(Ref.)	(Ref.)
Selfish Preferences				2.328***	2.732***
				(4.56)	(4.66)
Risk seeking				0.626*	0.953**
				(2.47)	(3.19)
Male				(Ref.)	(Ref.)
Female				−0.683	−0.700
				(−1.42)	(−1.36)
Perceived Competitiveness					−0.161
					(−1.48)
Perceived Competence					−0.196
					(−1.90)
Solved problems					−0.0296
					(−0.49)
Constant	−0.788*	−0.798*	−0.788*	−1.662	0.803
	(−2.53)	(−2.51)	(−2.53)	(−1.54)	(0.50)
N	144	144	144	144	144

z statistics in parentheses.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

individuals are supposed to be overconfident with a lower probability, as indicated also by the bivariate correlations in Table 1. The descriptive findings further reveal that selfish and hyperselfish decisions of overconfident leaders, i.e. hubris, are most often observed in the competitive treatment (overall 11 of 28 decisions). In the partial random treatment, we observe the fewest (hyper-)selfish decisions (overall 3 of 25 decisions). These findings are in line with our hypothesis and theory. In the pure random treatment, the amount of antisocial (hyper-)selfish decisions is located between both former treatments (overall 6 of 17 decisions) suggesting that, compared to the partly random treatment, some (low performing) leaders abuse their luck for their own benefit.

To rule out alternative explanations for the main findings, Table 3 reports the results of a regression analysis predicting antisocial decisions of leaders. Similar to Bendahan et al. (2015), we coded selfish decisions (option 3) and hyperselfish decisions (option 4) as 1, indicating abuse of power, and coded standard default decisions (option 2) and pro-social decisions (option 1) as 0. Model 1 and 2 show that the treatment and overconfidence do not affect dictator decisions per se although overconfidence is consistently, and in some of the model specifications significantly, positively correlated with selfish decisions in tendency. Model 3 investigate whether there is a significant interaction between overconfidence scores and selection method. Compared to the competitive treatment, overconfident leaders take significantly fewer anti-social decisions in the partly random treatment (p < .01), whereas there is no statistically difference, but also a consistent negative tendency, in the random treatment. Model 4 and 5 tests whether these effects remain robust and stable when controlling for alternative explanations for leaders' abuse of power, for example covariates like leaders' selfishness as well as endogenous covariates like perceived competence. The results remain stable.

Fig. 3 illustrated the main results of Model 3 in Table 3. The findings support the former descriptive results. The graph illustrates a significant difference between the selfishness of overconfident leaders in the partly random and the competitive treatment. In the partly random treatment only 10.2% of all overconfident leaders make antisocial decisions (p < .10). In the competitive treatment 42.3% of all overconfident leaders make antisocial decisions (p < .001). The difference is economically significant: in the partly random treatment only 1 of 10 overconfident leaders makes an antisocial decision. In the competitive treatment 4 of 10 overconfident leaders, i.e. 30% more, make an antisocial decision. It supports our hypothesis that overconfident leaders are not confirmed in their belief that they are far above the average and do not claim a major part of the pie when they experience that luck plays a role in being selected.

Antisocial decisions of leaders in the random treatment are located between both treatments with 32.4% (p < .01). As explained in the formal model, the lack of a performance screen in the random treatment has undesirable negative effects for community wealth, even if self-serving bias and fundamental attribution biases are triggered less. For the remaining underconfident leaders we find inconclusive patterns, but their behavior is beyond our research question.[13]

[13] With underconfident leaders, hubris plays no role, because according to our definition only overconfident people can display hubris. However, it would be interesting to study their behavior in the future. In the random treatment, the absolute number of underconfident leaders (31 leaders out of 48 leaders) is nearly twice as high as in the competition treatment (20 underconfident leaders out of 48 leaders) and in the partly random treatment (23 underconfident leaders out of 48 leaders), due to the missing competence screen. For underconfident group leaders, we find the opposite pattern as for overconfident group leaders. Underconfident leaders tend to abuse power more often in the partial

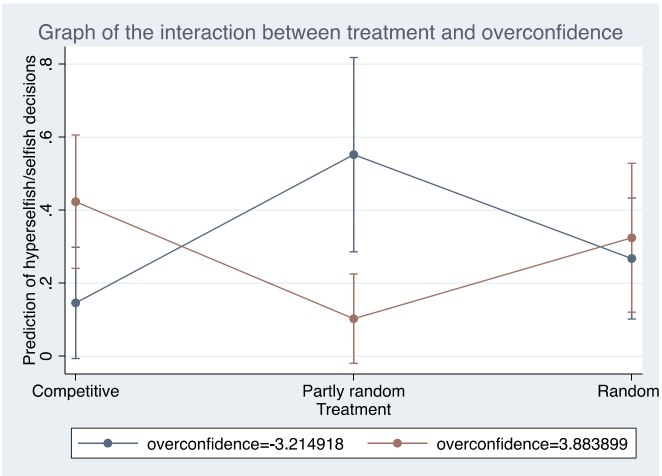

Fig. 3. Marginsplot of the interaction between treatment and overconfidence to predict hyperselfish and selfish decisions of leaders (estimations of Table 3, Model 3). Selfish decisions mean that the leader decided for Option 3 (for the leader: 270; for each follower: 130). Hyperselfish decisions mean that the leader decided for Option 4 (for the leader: 370; for each follower: 10). High overconfidence of leaders, the red line, is measured as the mean of overconfidence plus one standard deviation of overconfidence (3.883899). Low overconfidence of leaders, the blue line, is measured as the mean of overconfidence minus one standard deviation of overconfidence (−3.214918). The graph illustrates that overconfident leaders behave significantly less selfish in the partly random treatment as compared to the competitive treatment. The graph also illustrates that in the partly random treatment and to a lower degree also in the competitive treatment overconfident leaders show different degrees of selfish behavior as compared to underconfident leaders. It supports that both treatments trigger the hubris of leaders in contrariwise directions. The random treatment has no effects on the hubris of leaders; under- and overconfident leaders show the same degree of selfishness. The graph further indicates that, vice versa, underconfident leaders behave significantly more selfish in the partly random treatment as compared to the competitive treatment. Leaders underconfidence is not an essential part of our theory and research question and should be investigated by further research. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

As a robustness check, we tested whether the results remain stable in other estimation strategies. We estimated another binary logit model with a different coding of the dependent variable (1: hyperselfish, 0 selfish/standard/pro-social decision) and an ordered logit model with the full dependent variable (0 = pro-social, 1 = standard, 2 = selfish, 3 = hyperselfish). Table A3 and A4 in the appendix indicate that our main results remain stable across diverse methods of data analysis. Fig. A1 graphically illustrates the results of the interaction in Model 3 of Table A1. The results are comparable with the results in Fig. 3.

Accounting for competence

Our theory predicts that overconfident leaders in the competitive selection treatment will exhibit more hubris than in the two random selection treatments. Empirically, we find that over-confident leaders in the partly random selection treatment claim significantly higher compensations and managers in the random treatment do so too, at least in tendency. We also have shown, that these effects remain stable when controlling for the distributional norms prevailing in a group.

However, when we measured these norms, the subjects were not aware of the average performance difference between group leaders and group members in their treatment. As can be seen from Fig. 2, competence score differences are pronounced in the competitive selection treatment and virtually absent in the random selection treatment with the partly random treatment in between.

As such, it might well be the case that the group members, would they have been made aware of these differences, would have evaluated the compensation of the leaders differently. More to the point, building on equity theory (Adams, 1965), one could argue that leaders in the competitive treatment have performed better, hence it is fair for them to claim more. This view could be shared by a majority of followers in this treatment and in this case, the leaders' behavior would be in line with prevailing social norms and not hubristic at all.

To rule out this possibility, we conducted a further incentivized experiment. We invited subjects from the same pool to the lab and asked them which group leader compensation would be the most adequate, given the treatment-specific selection procedure and average competence score of the leaders and the followers in this treatment. The subjects (N = 168, 61.8% female) were randomly allocated to one of the three treatment conditions (competitive, random, partly random) and they only learned about the selection procedure and test scores in their specific treatment, but not in the other treatments. They then gave their opinion on which of the four options was the most adequate split. Participants on average gained USD 32 for 45 min (show-up fee of 15 USD and maximum of profit of 32 USD).

Table A5 in the appendix details the results of the second experiment, contrasting them with the actual behavior of all group leaders and the over-confident group leaders (i.e. those with an overconfidence score above zero). Here, we focus on the overconfident group leaders, because our theory predicts, that a random component in leader selection dampens hubris specifically among the over-confident.

In our second experiment, a majority of 55.4% evaluated the standard option as most adequate in the competitive treatment whereas in both treatments with a random component the most adequate option shifts to pro-social. However, this shift is rather small, with 52.63% choosing pro-social in the random treatment and 50.91% in the partly random treatment. This means that, despite of considerable variation in leader performance across treatments, normative expectations about adequate leader compensation differ only slightly across treatments.

Even more importantly, we do not find that the selfish or even hyperselfish option are viewed as adequate by a substantial share of subjects in the competitive selection treatment. On the contrary, both selfish options taken together are only viewed as adequate by a small minority of individuals in any treatment, namely by 7.14% (4 cases) in the competitive treatment, by 10.53% (6 cases) in the random treatment and by 1.82% (1 case) in the partly random treatment. These differences are not statistically significant ($\chi^2 = 3.52$, Cramér's V = 0.14, p = .172 across all three treatments and $\chi^2 = 1.83$, Cramér's V = 0.13, p = .176 when comparing competitive selection and partly random selection only).

This finding contrasts with the actual leader decisions in our behavioral experiment. As discussed in the main analyses, we find systematic differences in the allocation decisions of overconfident managers between the partly random treatment (12% selfish/hyperselfish, 3 cases) and the two other treatments (39.29% or 11 cases in the competitive treatment and 35.29% or 6 cases in the random treatment, chi squared test across all three treatments $\chi^2 = 5.316$, Cramér's V = 0.28, p = .07 or $\chi^2 = 5.06$, Cramér's V = 0.31, p = .025 when comparing competitive selection and partly random selection only).

The fact that we only find minor differences in normative expectations about leader behavior across treatments even when accounting for treatment-specific variation in leader performance, whereas we find more substantial and statistically significant treatment-specific variation in selfish leader-behavior rather supports the hubris theory than equity theory. Overconfident leaders do not hesitate to abuse their power for their own benefit, disregarding social norms under

(*footnote continued*)
random treatment (55.2% anti-social decisions) than in the competence treatment (14.6% anti-social decisions) or the random treatment (26.7% anti-social decisions).

performance selection, whereas hubristic behavior is dampened in the partly random selection treatment.

Discussion

Leadership hubris is a severe problem in many organizations (Hayward & Hambrick, 1997a; Hayward et al., 2006; Hiller & Hambrick, 2005; Roll, 1986). Leaders affected by hubris tend to overlook their limitations and take decisions that are harmful to the community (Billett & Qian, 2008; Hayward & Hambrick, 1997a; Malmendier & Tate, 2008). Previous studies on how governance mechanisms can prevent leadership hubris have mainly focused on board vigilance. These studies found that weak boards create more opportunities for hubristic leader decisions (Hayward & Hambrick, 1997b). Our research addresses one possible root of this problem and proposes an unusual selection practice for appointing leaders that mitigates hubris. To our knowledge, this is the first study that examines how organizations can tackle leadership hubris by applying particular selection practices to appoint their leaders. Drawing on historical evidence, we show that partly random selection, which combines competitive selection and random selection, helps organizations to reduce hubris.

We conducted a laboratory experiment whose results provide empirical evidence that this selection practice indeed reduces hubris in leaders. We found that with partly random selection, overconfident leaders claimed less for themselves and allocated more money to their subordinates. In this situation, overconfident leaders were less prone to misusing their power and took decisions that were more beneficial to the other members of the group than did overconfident leaders selected through competitive selection. Overconfident leaders selected by competitive screening tended to abuse their power, claiming large shares of the pie.

Theoretical implications

By adopting the idea of rational randomization, we contribute to the leadership literature, particularly to the literature on recruitment of leaders. The idea of partly random selection of leaders offers a novel perspective that contrasts sharply with suppositions that randomness is always irrational and harmful. However, seemingly rational decisions are often marred by many biases (Kahnemann, 2011). In such cases, the rationality of decision processes is a façade, and an intentionally random decision may be more rational.

We first show in a theoretical model that under realistic assumptions common welfare is higher under partly random selection than under competitive selection. Our formal model shows a tradeoff between the competence of selected candidates and their abuse of power. Organizations need both competent leaders and leaders that are not corrupt. Unfortunately, competent leaders selected for their performance tend to become overconfident over time. This development is also strengthened by Matthew effects, meaning that “the rich become richer and the poor become poorer” (Merton, 1968). Our model indicates that the disadvantage of less-competent leaders selected by lot can be compensated by the benefits of less corrupt behavior. The advantage of honest behavior is the greater the better the pre-selection of potential leaders according to competence and the nearer the candidates are to each other in competence.

We secondly show empirically that partly random selection avoids hubris in leaders. Overconfident leaders selected partly randomly are less prone to misusing their power. They take decisions that are more beneficial to other members of the group than do overconfident leaders selected through competitive selection.

Taking the findings together we enrich leadership theory, which has dealt intensively with leadership personalities and leadership styles (see e.g. Zhang, Ou, Tsui, & Wang, 2017) with a pioneering perspective that introduces partly random selection as a governance mechanism.

Practical implications

Our results have important practical implications. First, they enrich the toolkit of leadership recruitment with a novel instrument. We expect that this instrument can be applied as successfully as it was in Ancient Greece and in the Republic of Venice to prevent hubris and corruption. In contrast to many of the recommendations based on observation research, our practical implications are immune to endogeneity bias, an aspect often ignored (Antoniakis, Bendahan, Jaquart, & Lalive, 2010).

Second, our results can be applied not only with CEOs, but in very different fields of governance. It has been suggested to use random selection in Corporate Governance to enable a broader participation of stakeholders (see Zeitoun et al., 2014). Hubris of heads of state, who are convinced of their uniqueness and consider themselves to be above the normal rule of law (Woodruff, 2005; Einarsen, Aasland, & Skogstad, 2007; Owen & Davidson, 2009), can be mitigated by partly random selection (Duxbury, 2002). Citizen assemblies can be selected randomly (Fournier, 2011) and it has been suggested to introduce a second chamber to the EU parliament selected by lot (Buchstein, 2009; Frey & Osterloh, 2016). The idea that random selection should have a place in modern political arrangements has received considerable support among contemporary political theorists (e.g. Buchstein, 2010; Duxbury, 2002; Manin, 1997; Sintomer, 2014).

Third, recruitment consultants will not lose their jobs. We have demonstrated theoretically that the better they work in selecting a pool of candidates, the more important is the advantage of partly random selection in dampening hubris.

Limitations and suggestions for future research

Our study has several limitations that open opportunities for further research.

First, the experimental finding that partly random selection dampens leadership hubris cannot be generalized directly to real-life settings (Levitt & List, 2007). Most importantly, we do not know whether leaders in organizations react similarly to students. Whereas laboratory experiments are well suited to investigating how institutions shape human behavior (Guala, 2005), and students in the laboratory act similarly to the general population concerning prosocial behavior (Benz & Meier, 2008), we do not know whether this is also the case in this specific situation (Falk, Meier, & Zehnder, 2013). Indeed, Brunell et al. (2008) find that CEOs in general are more narcissistic, and thus more prone to hubris, than students. This finding suggests that partly random selections could be even more effective among real-life leaders.

Second, the generalizability of our findings is limited because of the different time scales of experimental and real-life settings. The duration of our experiment was very short, whereas in organizations the process through which CEOs are appointed is much longer. Moreover, candidates in the experiment had to answer simple questions, whereas search committees in organizations have to screen and compare different, often difficult-to-measure qualities of candidates.

Third, our experiment did not investigate competence and hubris effects in an integrated approach that allowed common welfare to be measured more directly. As the formal model suggests, such a test would be useful for investigating the trade-offs between the costs of abusing power and the costs of lacking competence. It could also be used to investigate in more detail the parameters of these trade-offs in a range of selection methods.

Fourth, the pools of candidates in our experiment consisted only of groups of six individuals, which leads to small differences in average competence between those selected in purely random selections and in partly random selections. Larger pools of candidates may increase performance differences between candidates in both treatments and may also enlarge differences in leadership hubris effects between the selection methods.

Fifth, competitive selections not only strengthen the hubris of leaders but may also lead to an overselection of overconfident people into powerful positions. In our empirical design, we were able to test the first effect but not the second. Further research is needed to study whether overconfident people are more often considered to be effective leaders in competitive selections, and whether consequently more leaders are inclined to hubris.

Sixth, it has been argued that to enable an innovative culture, the narcissism and humility of leaders have to interact (Zhang et al., 2017). Also, with the interaction of hubris and prosocial behavior could be important in fostering innovation. An experiment could be designed to explore whether random selection has an impact on innovativeness.

Seventh, our findings about the social behavior of underconfident leaders were inconclusive, but their behavior is beyond our research question. Nevertheless, it would be interesting to study this behavior in the future.

Finally, there is much room for further research with field studies. One could conduct a field experiment with teamwork between students that applied random selection of the team leader. Comparing the behavior of the team leader and the team members with conventional teamwork could produce valuable insights into the effects of random selection. In the future, observation studies could also be employed, for instance if in Switzerland the popular initiative to draw federal judges out of a pool is implemented.

Conclusion

Our study follows a pioneering approach to investigate an unusual selection method for appointing leaders in organizations, partly random selection. This selection method has been extensively used in history but has nearly been forgotten. Today, random decisions are considered by many people to be “irrational”. Our study shows that purposeful random selection, in particular combining competitive selections with a random component, is a rational and promising way of recruiting leaders that tackles hubris in overconfident leaders. Our proposal to “draw your CEO by lot” is provocative but may be promising.

Appendix 1

Table A1
Scores in the competence test measured by the number of solved tasks in the different treatments.

All subjects	Mean	Std. Err.	Unadjusted [95% Conf. Interval]	
Competitive treatment	9.607639	0.270365	9.076987	10.13829
Random treatment	9.944444	0.270365	9.413793	10.4751
Partly random treatment	9.083333	0.270365	8.552682	9.613985
N	864	864	864	864

All subjects	Contrast	Std. Err.	Unadjusted	
			t	P > |t|
Random treatment vs Competitive treatment	0.3368056	0.3823538	0.88	0.379
Partly random treatment vs Competitive treatment	−0.5243056	0.3823538	−1.37	0.171
Partly random treatment vs Random treatment	−0.8611111* **	0.3823538	−2.25	0.025
N	864	864	864	864

Only group leaders	Mean	Std. Err.	Unadjusted [95% Conf. Interval]	
Competitive treatment	15.54167	0.5792505	14.39653	16.68681
Random treatment	11.77083	0.5792505	10.62569	12.91597
Partly random treatment	12.10417	0.5792505	10.95903	13.24931
N	144	144	144	144

Only group leaders	Contrast	Std. Err.	Unadjusted	
			t	P > |t|
Random treatment vs Competitive treatment	−3.770833***	0.819184	−4.60	0.000
Partly random treatment vs Competitive treatment	−3.4375***	0.819184	−4.20	0.000
Partly random treatment vs Random treatment	0.3333333	0.819184	−0.41	0.685
N	144	144	144	144

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

The table illustrates the competence of participants and group leaders across treatment conditions. Competence is measured as the number of correctly solved task out of 30 tasks in a standardized general knowledge test under time pressure.

Table A2
Descriptive statistic of the number of overconfident managers in the different treatments and their selfish decision making.

Overconfident managers (overprediction of the realized competence score)	Competitive treatment	Random treatment	Partly random treatment
Pro-social decisions	12	8	18
Standard decisions	5	3	4
Selfish decisions	3	1	2
Hyperselfish decisions	8	5	1
N	28	17	25
% of all managers in the treatment	58%	35%	52%

Standard decisions mean that the leader in the splitting decision decided for Option 1 (for the leader: 220 MP; for each follower: 190 MP). Equal decisions mean that the leader decided for Option 2 (for the leader: 210 MP; for each follower: 210 MP). Selfish decisions mean that the leader decided for Option 3 (for the leader: 270; for each follower: 130). Hyperselfish decisions mean that the leader decided for Option 4 (for the leader: 370; for each follower: 10).

Table A3
Binary logit regression models to predict hyperselfish decisions of leaders (1: hyperselfish, 0 selfish/standard/pro-social decision).

	Model 1	Model 2	Model 3	Model 4	Model 5
Competitive treatment	(Ref.)	(Ref.)	(Ref.)	(Ref.)	(Ref.)
Random treatment	0.131	0.186	1.167	1.122	−0.632
	(0.26)	(0.36)	(1.49)	(1.18)	(−0.43)
Partly random treatment	0.131	0.134	1.186	1.127	1.221
	(0.26)	(0.26)	(1.47)	(1.15)	(0.97)
Overconfidence		0.042	0.426**	0.509*	0.658*
		(0.68)	(2.59)	(2.35)	(2.19)
Competitive treatment × Overconfidence			(Ref.)	(Ref.)	(Ref.)
Random treatment × Overconfidence			−0.356	−0.467	−0.522
			(−1.85)	(−1.89)	(−1.53)
Partly random treatment × Overconfidence			−0.845***	−0.941**	−1.286**
			(−3.53)	(−3.03)	(−2.96)
Prosocial Groupnorm				0.823	0.462
				(0.75)	(0.37)
Selfish Preferences				2.802***	4.168***
				(4.35)	(4.50)
Risk seeking				0.907***	1.756***
				(2.91)	(3.72)
Male				(Ref.)	(Ref.)
Female				−0.386	−0.136
				(−0.69)	(−0.20)
Perceived Competitiveness					−0.377*
					(−2.46)
Perceived Competence					−0.340*
					(−2.03)
Solved problems					0.130
					(1.52)
Constant	−1.466***	−1.519***	−2.506***	−7.393**	−7.648**
	(−3.97)	(−3.99)	(−3.60)	(−4.08)	(−2.68)
N	144	144	144	144	144

z statistics in parentheses.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

Table A4
Ordered logit regression models to predict selfish decisions of leaders (0 = pro-social, 1 = standard, 2 = selfish, 3 = hyperselfish).

	Model 1	Model 2	Model 3	Model 4	Model 5
Competitive treatment	(Ref.)	(Ref.)	(Ref.)	(Ref.)	(Ref.)
Random treatment	−0.274	−0.121	−0.121	−0.079	−0.644
	(−0.68)	(−0.29)	(−0.29)	(−0.16)	(−1.07)
Partly random treatment	−0.227	0.148	0.148	0.103	−0.181
	(−0.57)	(0.35)	(0.35)	(0.21)	(−0.34)
Overconfidence		0.030	0.156	0.097	0.103
		(0.61)	(1.85)	(1.01)	(1.07)
Competitive treatment × Overconfidence			(Ref.)	(Ref.)	(Ref.)
Random treatment × Overconfidence			−0.069	−0.051	−0.049
			(−0.56)	(−0.37)	(−0.36)
Partly random treatment × Overconfidence			−0.369**	−0.291	−0.320*
			(−2.74)	(−1.94)	(−2.14)
Prosocial Groupnorm				−1.421**	−1.505**
				(−2.55)	(−2.64)
Selfish Preferences				2.200***	2.349***
				(5.44)	(5.55)
Risk seeking				0.627**	0.726**
				(2.96)	(3.24)
Male				(Ref.)	(Ref.)
Female				−0.245	−0.275
				(−0.62)	(−0.68)
Perceived Competitiveness					−0.162
					(−1.85)
Perceived Competence					−0.0183
					(−0.23)
Solved problems					−0.0119
					(−0.25)
cut1					
_cons	0.262	0.287	0.405	1.705	0.388
	(0.95)	(1.03)	(1.35)	(1.900)	(0.31)
cut2					
_cons	0.729**	0.754**	0.890**	2.372**	1.071
	(2.59)	(2.65)	(2.91)	(2.62)	(0.85)
cut3					
_cons	1.220***	1.245***	1.415***	3.073***	1.811
	(4.10)	(4.15)	(4.36)	(3.318)	(1.43)
N	144	144	144	144	144

z statistics in parentheses.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

Table A5
Per cent of subjects choosing the respective option as the most adequate, given the selection rule, average test score of group leaders and followers (first row of each cell; majority vote bold), actual decisions of the group leaders (second row) and actual decisions of the overconfident group leaders (third row).

	Competitive treatment	Random treatment	Partly random treatment
Pro-social decisions	37.50	**52.63**	**50.91**
	54.17	64.58	62.50
	42.86	47.06	72.00
Standard decisions	**55.36**	36.84	47.27
	14.58	6.25	10.42
	17.86	17.65	16.00
Selfish decisions	7.14	3.51	1.82
	12.50	8.33	6.25
	10.71	5.88	8.00
Hyperselfish decisions	0.0	7.02	0.0
	18.75	20.82	20.83
	28.57	29.41	4.00

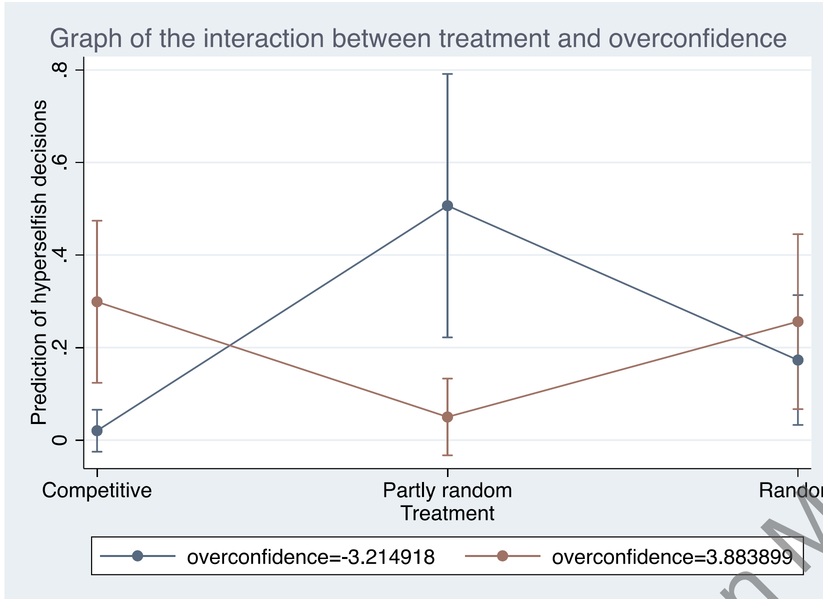

Fig. A1. Marginsplot of the interaction between treatment and overconfidence to predict hyperselfish decisions of leaders (estimations of Table A1, Model 3). Hyperselfish decisions mean that the leader decided for Option 4 (for the leader: 370; for each follower: 10). High overconfidence of leaders, the red line, is measured as the mean of overconfidence plus one standard deviation of overconfidence (3.883899). Low overconfidence of leaders, the blue line, is measured as the mean of overconfidence minus one standard deviation of overconfidence (−3.214918). The graph illustrates that overconfident leaders behave significantly less hyperselfish in the partly random treatment as compared to the competitive treatment. The graph also illustrates that in the partly random treatment and in the competitive treatment overconfident leaders show different degrees of hyperselfish behavior as compared to underconfident leaders. It supports that both treatments trigger the hubris of leaders in contrariwise directions. The random treatment has no effects on the hubris of leaders; under- and overconfident leaders show the same degree of hyperselfishness. The graph further indicates that, vice versa, underconfident leaders behave significantly more hyperselfish in the partly random treatment as compared to the competitive treatment. Leaders underconfidence is not an essential part of our theory and research question and should be investigated by further research.

Appendix 2

Formal account of competitive selections as a trigger for leadership hubris

To better understand the mechanism through which hubris occurs, we provide a more formal, descriptive account of the economic effects of selection methods on the common welfare of the group in question. The analysis focuses on the tradeoffs between the gross positive competence effects of competitive selections and the positive effects of preventing the misuse of power by partly random selections. We model these determinants to establish a basic framework for the experimental work that is presented below.[14]

In general, we assume that groups screen leaders before hiring them to secure common welfare, for example the welfare of an organization, firm, or state. We define welfare as the difference between advantages of high leadership productivity gained by rigorous selection minus the costs of misuse of power, for example due to hubris, and an average market wage W. It implies searching for highly competent leaders and avoiding the costs of misuse of power. For simplicity, we assume the pool of leaders in the population consists of types with either *high* or *low competence* (H or L) which are distributed in the population with probability p for H and $(1 - p)$ for L. They also exhibit dispositions for either *good* or *bad behavior* (G or B), leading eventually to the abuse of power. We assume that these leader types are uniformly distributed within the population of P_H - and P_L-type leaders. Let the gross productivity of the four types in the population be

$$P_{IJ} \varepsilon \{(H,G);(H,B);(L,G);(L,B)\}. \tag{A1}$$

For the gross productivity of the leaders' net of abuse of power costs we assume:

$$P_{HGG} > P_{HBG} > W > P_{LGG} > P_{LBG} \text{ with average Wage } W. \tag{A2a}$$

and the zero profit condition:

$$(P_{HGG} + P_{HBG} + P_{LGG} + P_{LBG}) - 4W = 0. \tag{A2b}$$

We define the net productivity of the different types of leaders as: $P_{HG} = P_{HGG} - W$,

$$P_{HB} = P_{HBG} - W,\ P_{LG} = P_{LGG} - W \text{ and } P_{LB} = P_{LBG} - W. \tag{A2c}$$

We assume that a participation constraint $W > U_{ij}$ (A2d) holds for any type of leader. This constraint will never be binding in our model which follows immediately from our assumptions both that leaders do not incur costs for the screening nor have to exert any work effort.

To maximize its welfare a firm has to solve a two step selection problem: first because of $P_{H,j} > P_{L,j}\ \forall j$ it has to find the H-type managers, then in the second step out of this pool the G-Type managers need to be selected. This task gets complicated by information problems about productivity and managerial behavior the firm faces.

We hypothesize that a leader with strong overconfidence who went through a competitive selection process will act corruptly and vice versa. To model

[14] In our experiment we concentrate on the separate analysis of competence on one hand and hubris on the other hand. We did not integrate both factors into one aggregate measure of common welfare.

the consequences of this hypothesis, we compare the common welfare of the group in three selection processes: purely random (with zero screening costs), which is used as a point of reference, competitive selection, and partly random selection. The costs of screening are $0 < s$ and are assumed to be identical for the two latter selection processes. The quality of screen is $0 < q < 1$ with probability of correct decision q and probability of wrong decision $(1 - q)$.

Common welfare of purely random selection versus *competitive selection.*

The expected gross welfare of the firm, defined as the expected value of net leader productivity, in the pure random case is

$$\pi_R = (p/2\ P_{HG} + p/2\ P_{HB}) + (1/2(1-p)\ P_{LG} + (1/2\ (1-p)\ P_{LB}) \quad \text{(A3)}$$

The lack of a performance screen in a pure random selection will have unavoidable negative productivity effects that in most cases will not be overcompensated by positive (anti) "hubris" effects.[15]

We conjecture that the more competitive the leader's selection process is, the more overconfident leaders abuse their power, and in that respect, produce costs for the firm. The selection process can be viewed as a trigger for this sort of behavior, which either hides or brings to surface the underlying dispositions. The trigger is the information to the leader how he was selected. The trigger is modeled with $(1 - t)$ with $0 < t < 1$, implying that a high (low) trigger $(1 - t)$ has a strong (small) impact on the willingness of highly (and lowly) productive leaders to act corruptly compared to the pure random selection; the effect for P_L-leaders will be the smaller the better q is.

For the competitive selection we get for the expected common welfare net of screening costs:

$$\pi_{SC} = qtpP_{HG} + q(1-t)pP_{HB} + (1-q)t(1-p)\ P_{LG} + (1-q)(1-t)\ (1-p)P_{LB} - s_{SC}. \quad \text{(A4a)}$$

with π_{SCG} as gross welfare and π_{SC} as welfare net of screening costs and with

$$s_{SC} = \alpha\pi_{SCG} \text{ with } 0 < \alpha < 1 \text{ and } \pi_{SC} = (1-\alpha)\ \pi_{SCG} \quad \text{(A4b)}$$

It holds that $\frac{\partial\pi_{SC}}{\partial_S} < 0,\ \frac{\partial\pi_{SC}}{\partial_t} > 0,\ \frac{\partial\pi_{SC}}{\partial_{(1-t)}} < 0$ (A5)

The quality of the screen assumes that a fraction $(1 - q)$ of P_H-leaders (*high competence*) are mistakenly rejected, and the same fraction of P_L-leaders (*low competence*) are wrongly chosen. Dispositions for bad behavior that some leaders have might be brought to surface by the trigger $(1 - t)$, the information that they are winners of a competitive selection process. We assume that the trigger works depending on the distribution of a psychic disposition in the population of leaders. Given a high quality of screening, it follows for a high fraction of P_H-leaders that their dispositions for bad behavior are triggered. Given that $P_{HG} > P_{HB} > W > P_{LG} > P_{LB}$, we can compare the outcomes of pure random and competitive selection. It is obvious that, ceteris paribus, competitive selection with a relatively high q will bring down the selection of unproductive leaders and will therefore increase common welfare compared to pure randomness. Higher t in the population of leaders helps to increase common welfare by increasing the marginal probability for choosing uncorrupt leaders, who are either productive or unproductive.

Proposition 1

Comparing common welfare of the competitive selection with pure random selection, we have to note that only q can be "deliberately" chosen by the firm to improve its common welfare. For $P_{HG} > P_{HB} > W > P_{LG} > P_{LB}$ (A2a) with average Wage W and the zero profit condition for pure random selection: $(P_{HG} + P_{HB} + P_{LG} + P_{LB}) - 4\ W = 0$ (A2b) there always exists at least one q^* which satisfies the net profit condition

$$(1-\alpha)\pi_{SCG} = \pi_{SC} > \pi_R = 0. \quad \text{(A6)}$$

Therefore, competitive selection always dominates pure random selection.

Proof

Comparing common welfare of the competitive selection with pure random selection, we start with the fairly general assumption of $P_{HG} > P_{HB} > W > P_{LG} > P_{LB}$ (A2a) with average Wage W and the zero profit condition for pure random selection $(P_{HG} + P_{HB} + P_{LG} + P_{LB}) - 4\ W = 0$ (A2b). This latter condition is the expected zero profit of pure random selection.

Therefore to show under which conditions $\pi_{SC} > \pi_R$ holds we have to proof that there exits at least one q^* that leads both to a positive gross profit $\pi_{SCG} > 0$ and to a positive net profit $\pi_{SC} = \pi_{SCG} - s_{SC} > 0$.

For ease of comparison we start with the assumption: $t = 0{,}5$ and $p = 0{,}5$, which implies for the quality of the screening $q = 1/2$ *that:* $\pi_{SC} = \pi_{RS} = 0$. It follows

$\pi_{RS} = p/2\ P_{HG} + p/2\ P_{HB} + 1/2(1-p)\ P_{LG} + 1/2\ (1-p)\ P_{LB}$ and

$\pi_{SC} = qtpP_{HG} + q(1-t)pP_{HB} + (1-q)t(1-p)\ P_{LG} + (1-q)(1-t)\ (1-p)P_{LB} - s_{SC}$.

For convenience we simplify the notation and define (A7):

(A7a) $A = P_{HG}$

(A7b) $B = P_{HB}$

(A7c) $C = P_{LG}$

(A7d) $D = P_{LB}$

It follows immediately:

$$\pi_R = 1/4\ P_{HG} + 1/4\ P_{HB} + 1/4\ P_{LG} + 1/4\ P_{LB}. \quad \text{(A8a)}$$

Inserting (A(7a)) to (A(7d)) leads to:

$$\pi_R = A/4 + B/4 + C/4 + D/4 \text{ and: (A8c) } \pi_{SCG} = qA/4 + qB/4 + (1-q)C/4 + (1-q)D/4 \quad \text{(A8b)}$$

So $\pi_{SCG} > \pi_R$ holds for every q^* that both satisfies

[15] Using a pure random selection process means either finding a highly productive leader P_H with probability p or ending up with an unproductive leader P_L with probability (1–p), depending on the population of leaders. Of course, both leader groups consist of either good or bad types of leaders. It becomes obvious that when there is even a very small excess of low over high productivity leaders in the population, then using the usual assumption for screening models $P_{HG} > P_{HB} > W > P_{LG} > P_{LB}$ will inevitably lead to an expected negative common welfare in this selection process; therefore pure random selection can't be the optimal solution in most practical cases.

$$\frac{q}{(q-1)} > \frac{(A+B)}{(C+D)} = -1 \tag{A9}$$

This holds because of (A7a) and (A7b): $(A + B) > 0$ stands for the negative effects of not choosing highly productive managers and because of (A7c) and (A7d) leads to: $(C + D) < 0$ which stands for the avoided wrong hiring decisions of lowly productive managers.

(A9) implies for $t = 0{,}5$ and $p = 0{,}5$ and $s_{SC} > 0$ that there exists always a q^* *for all* $q \in \left(\frac{1}{2}, 1\right)$

Given that (A5) $\frac{\partial \pi_{SC}}{\partial s} < 0$, $\frac{\partial \pi_{SC}}{\partial t} > 0$, $\frac{\partial \pi_{SC}}{\partial (1-t)} < 0$ it follows immediately that even for the "worst case" parameter constellation for competitive selection which is:

$$0 < p < ½ \tag{A10a}$$

which implies that p is below the assumed probability $p = ½$ in the relevant reference case of partly random selection - and

$$0 < t < ½ \tag{A10b}$$

there is always a q* with: $½ < q^* < 1$ which satisfies $\pi_{SC} > 0 > \pi_R = 0$. The quality of the screen q^* can even fall below ½: (A10c) $q^* < ½$ *for* $p > ½$ and $t > ½$ to satisfy $\pi_{SC} > 0 > \pi_R = 0$.

Therefore, competitive selection *always* dominates pure random selection.

Common welfare of competitive selection versus *partly random selection*

Partly random selection implies that P_L-leaders (*low competence*) have a better chance of being selected than in the competitive selection case. This fact is expressed by r for the change in probability to be chosen as a P_L-type with $0 < r < 1$. Compared to the competitive case, a high r can lead to the selection of more of the lowly productive leaders who potentially abuse their power. But using r also completely eliminates the trigger $(1 - t)$, because the winners now come out of a partly random selection process. The expected common welfare in the partly random selection case becomes

$$\pi_{RS} = \{(1-r)qp\,P_{HG} + q(1-r)pP_{HB}\} + \{(1-q)r(1-p)P_{LG} + (1-q)r(1-p)P_{LB}\} - s_{RS} \tag{A11a}$$

with π_{RSG} as gross welfare and π_{RS} as welfare net of screening costs and with $s = \alpha\,\pi_{RS}$ *with* $0 < \alpha < 1$ and $\pi_{SC} = (1 - \alpha)\,\pi_{RSG}$

$$\text{It holds that: } \frac{\partial \pi_{RS}}{\partial s} < 0,\ \frac{\partial \pi_{RS}}{\partial r} < 0,\ \frac{\partial \pi_{RS}}{\partial (1-r)} > 0. \tag{A11b}$$

Comparing common welfare of the competitive selection with partly random selection, we have to note that t is an exogenously distributed predisposition of leaders in the population whereas r can be deliberately chosen by the firm to improve its common welfare.

Proposition 2

When $\pi_{RS} > \pi_{SC}$ for all t and r an appropriately designed partly random selection will always dominate pure competitive selection as a screening strategy. There exists an r^* with $1 > r^* > 0$ for every t with $1 \geq t > 0$ which satisfies: $\pi_{RS}(r^*) > \pi_{SC}(t)$

Proof

To show that there always exists an r^* with $1 > r^* > 0$ for every t with $1 \geq t > 0$ which satisfies:

$\pi_{RS}(r^*) > \pi_{SC}(t_{max})$ it suffices to demonstrate there always exists an r^* even for the *best* case (i.e. the one with the highest welfare) for competitive selection which is by definition: $t = 1$.

In order to do so we compare (A11a) and (A4a) for $t_{max} = 1$.

For convenience we again simplify the notation and define (A12):

(A12a) $A = P_{HG} > 1$, *(A12b)* $B = P_{HB} > 1$, *(A12c)* $C = P_{LG} < 0$, *(A12d)* $D = P_{LB} < 0$, *(A12e)*

$a = pq$ *with* $1 > a > 0$

(A12f) $b = (1-q)(1-p)$ *with* $1 > b > 0$; Given $A > B > 1 > 0 > C > D$ and using:

(A2a) $(A\text{-}D) > (B\text{-}C)$ and *(A2b)* $(A + B + C + D) = 0$

And assuming $\pi_R(r^*) > \pi_{SC}(t_{max})$ the comparison of (A11a) and (A4a) for $t = 1$ leads to:

$-raA + (1\text{-}r)aB + (r\text{-}1)\,bC + rbD > 0.$

Rearranging terms leads to:

$$(1-r)/r > (aA - bD)/(aB - bC) \tag{A13}$$

which equals:

$$\frac{1-r}{r} > (aP_{HG} - bP_{LB})/(aP_{HB} - b\,P_{LG}) \tag{A14}$$

So, because of *(A2a)*, *(A2b)*, *(A12e)* and *(A12f)* there always exists at least one r^* that satisfies (3).

Therefore, an appropriately designed partly random selection will *always* dominate pure competitive selection as a screening strategy.

This result can be illustrated with numerical examples. Of course, for high t the r^* that improves welfare has to become very small implying that the selection process of the firm basically approaches the competitive selection process in reality.

References

Adams, J. S. (1965). Inequity in social-exchange. *Advances in Experimental Social Psychology, 2*(4), 267–299.

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49*(6), 1621.

Antoniakis, J., Bendahan, S., Jaquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendation. *The Leadership Quarterly, 21*, 1086–1120.

Aristotle (2003). *Rhetoric.* Translated by J. H. FreeseOxford: Oxford University Press.

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics, 116*(1), 261–292.

Bartlett, M. Y., & DeSteno, D. (2006). Gratitude and prosocial behavior: Helping when it costs you. *Psychological Science, 17*(4), 319–325.

Bénabou, R., & Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *The Journal of Economic Perspectives, 30*(3), 141–165.

Bendahan, S., Zehnder, C., Pralong, F. P., & Antonakis, J. (2015). Leader corruption depends on power and testosterone. *The Leadership Quarterly, 26*(2), 101–122.

Benz, M., & Meier, S. (2008). Do people behave in experiments as in the field?—Evidence from donations. *Experimental Economics, 11*(3), 268–281.

Bertrand, M., & Mullainathan, S. (2001). Are CEOs rewarded for luck? The ones without

principals are. *The Quarterly Journal of Economics, 116*(3), 901–932.
Billett, M. T., & Qian, Y. (2008). Are overconfident CEOs born or made? Evidence of self-attribution bias from frequent acquirers. *Management Science, 54*(6), 1037–1051.
Blair, C. A., Hoffman, B. J., & Helland, K. R. (2008). Narcissism in organizations: A multisource appraisal reflects different perspectives. *Human Performance, 21*(3), 254–276.
Blickle, G., Schlegel, A., Fassbender, P., & Klein, U. (2006). Some personality correlates of business white-collar crime. *Applied Psychology, 55*(2), 220–233.
Boochs, W. (2004). *Geschichte und Geist der koptischen Kirche.* Bernardus-Verlag: Langwarden.
Brown, J. D. (1986). Evaluations of self and others: Self-enhancement biases in social judgments. *Social Cognition, 4*(4), 353–376.
Brunell, A. B., Gentry, W. A., Campbell, W. K., Hoffman, B. J., Kuhnert, K. W., & DeMarree, K. G. (2008). Leader emergence: The case of the narcissistic leader. *Personality and Social Psychology Bulletin, 34*(12), 1663–1676.
Buchstein, H. (2009). *Demokratie und Lotterie: Das Los als politisches Entscheidungsinstrument von der Antike bis zur EU.* (Frankfurt am Main/New York: Campus).
Buchstein, H. (2010). Reviving randomness for political rationality: Elements of a theory of aleatory democracy. *Constellations, 17*(3), 435–454.
Buchstein, H. (2019). Zufallsentscheidungen historisch betrachtet (Special Issue on "Leadership and Random Decision", edited by Osterloh M, Rost K). *Zeitschrift Führung und Organisation, 88*(3), 162–168.
Burckhardt, A. (1916). Über die Wahlart der Basler Professoren, besonders im 18. *Jahrhundert.Basler Zeitschrift für Geschichte und Altertumskunde, 15*, 28–46.
Cairns, D. L. (1996). Hybris, dishonour, and thinking big. *The Journal of Hellenic Studies, 116*, 1–32.
Caligor, E., Levy, K. N., & Yeomans, F. E. (2015). Narcissistic personality disorder: diagnostic and clinical challenges. *172*(5), 415–422.
Camerer, C., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *The American Economic Review, 89*(1), 306–318.
Chatterjee, A., & Hambrick, D. C. (2007). It's all about me: Narcissistic chief executive officers and their effects on company strategy and performance. *Administrative Science Quarterly, 52*, 352–386.
Dobbins, G. H., & Russell, J. M. (1986). Self-serving biases in leadership: A laboratory experiment. *Journal of Management, 12*(4), 475–483.
Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2011). Individual risk attitudes: Measurement, determinants and behavioral consequences. *Journal of the European Economic Association, 9*(3), 522–550.
Dowlen, O. (2017). *The political potential of sortition: A study of the random selection of citizens for public office.* Andrews UK Limited.
Duxbury, N. (2002). *Random justice: On lotteries and legal decision-making.* Oxford: Oxford University Press.
Einarsen, S., Aasland, M. S., & Skogstad, A. (2007). Destructive leadership behaviour: A definition and conceptual model. *Leadership Quarterly, 18*(3), 207–216.
Falk, A., Meier, S., & Zehnder, C. (2013). Do lab experiments misrepresent social preferences? The case of self-selected student samples. *Journal of the European Economic Association, 11*(4), 839–852.
Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178.
Fournier, P. (2011). *When citizens decide: Lessons from citizen assemblies on electoral reforms.* (New York).
Frank, R. H. (2016). *Success and luck: Good fortune and the myth of meritocracy.* Princeton University Press: Princeton.
Frey, B. S., & Osterloh, M. (2016). Aleatoric democracy. *CESifo Working Paper Nr. 6229*.
Frey, B. S., & Steiner, L. (2014). God does not play dice, but people should: Random selection in politics, science and society. *CREMA (Center for Research in Economics, Management and the Arts) working paper no. 2014-09*.
Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117*(1), 21.
Greif, A. (1995). Political organizations, social structure, and institutional success: Reflections from Genoa and Venice during the commercial revolution. *Journal of Institutional and Theoretical Economics (JITE),* 734–740.
Guala, F. (2005). *The methodology of experimental economics.* Cambridge University Press.
Hansen, M. H. (1991). *The Athenian democracy in the age of Demosthenes: Structure, principles, and ideology.* University of Oklahoma Press.
Hayward, M. L., & Hambrick, D. C. (1997a). Explaining the premiums paid for large acquisitions: Evidence of CEO hubris. *Administrative Science Quarterly,* 103–127.
Hayward, M. L. A., & Hambrick, D. C. (1997b). Explaining the premiums paid for large acquisitions: Evidence of CEO hubris. *Administrative Science Quarterly, 42*(1), 103–127.
Hayward, M. L. A., Shepherd, D. A., & Griffin, D. (2006). A hubris theory of entrepreneurship. *Management Science, 52*(2), 160–172.
Hiller, N. J., & Hambrick, D. C. (2005). Conceptualizing executive hubris: The role of (hyper-)core self-evaluations in strategic decision-making. *Strategic Management Journal, 26*(4), 297–319.
Hollow, M. (2014). The 1920 Farrow's Bank failure: A case of managerial hubris. *Journal of Management History, 20*(2), 164–178.
Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review, 76*(4), 728–741.
Kahnemann, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.
Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review, 110*(2), 265–284.
Lammers, J., Galinsky, A. D., Dubois, D., & Rucker, D. D. (2015). Power and morality. *Current Opinion in Psychology, 6*, 15–19.
Larkin, I., & Leider, S. (2012). Incentive schemes, sorting, and behavioral biases of employees: Experimental evidence. *American Economic Journal: Microeconomics, 4*(2), 184–214.
Larwood, L., & Whittaker, W. J. J. (1977). Managerial myopia: self-serving biases in organizational planning. *62*(2), 194.
Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives, 21*(2), 153–174.
Liu, C., & De Rond, M. (2016). Good night, and good luck: Perspectives on luck in management scholarship. *The Academy of Management Annals, 10*(1), 409–451.
Malmendier, U., & Tate, G. (2005). CEO overconfidence and corporate investment. *The Journal of Finance, 60*(6), 2661–2700.
Malmendier, U., & Tate, G. (2008). Who makes acquisitions? CEO overconfidence and the market's reaction. *Journal of Financial Economics, 89*(1), 20–43.
Malmendier, U., & Taylor, T. (2015). On the verges of overconfidence. *Journal of Economic Perspectives, 29*(4), 3–8.
Manin, B. (1997). *The principles of representative government.* Cambridge: Cambridge University Press.
McCormick, J. P. (2006). Contain the wealthy and patrol the magistrates: Restoring elite accountability to popular government. *American Political Science Review, 100*(2), 147–163.
Merton, R. K. (1968). The Matthew effect in science. *Science, 159*(3810), 56–63.
Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction. *Psychological Bulletin, 82*(2), 213–225.
Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502–517.
Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior, 19*(3), 338–368.
Nevicka, B., Ten Velden, F. S., De Hoogh, A. H., & Van Vianen, A. E. (2011). Reality at odds with perceptions: Narcissistic leaders and group performance. *Psychological Science, 22*(10), 1259–1264.
Nicolas, R., & Fisher, E. (1992). *Hybris: A Study in the Values of Honour and Shame in Ancient Greece.* Warminster: Aris & Phillips.
Obermayer, B., & Obermaier, F. (2016). *Panama papers. Die Geschichte einer weltweiten Enthüllung.* Köln: Kiepenheuer & Witsch.
O'Connor, J., Mumford, M. D., Clifton, T. C., Gessner, T. L., & Connelly, M. S. (1996). Charismatic leaders and destructiveness: An historiometric study. *The Leadership Quarterly, 6*(4), 529–555.
Osterloh M, Frey BS. Forthcoming. How to avoid borrowed plume in academia. Research Policy.
Oswald AJ Forthcoming. Rational randomization by journal editors. Research Policy.
Owen, D. (2012). *The hubris syndrome: Bush, blair and the intoxication of power.* London: Methuen.
Owen, D., & Davidson, J. J. B. (2009). Hubris syndrome: An acquired personality disorder? A study of US Presidents and UK Prime Ministers over the last 100 years. *Brain: A Journal of Neurology, 132*(5), 1396–1406.
Paunonen, S. V., Lönnqvist, J.-E., Verkasalo, M., Leikas, S., & Nissinen, V. (2006). Narcissism and emergent leadership in military cadets. *The Leadership Quarterly, 17*(5), 475–486.
Petit, V., & Bollaert, H. (2012). Flying too close to the sun? Hubris among CEOs and how to prevent it. *Journal of Business Ethics, 108*(3), 265–283.
Pocock, J. G. A. (2016). *The Machiavellian moment: Florentine political thought and the Atlantic republican tradition.* Princeton: Princeton University Press.
Roll, R. (1986). The hubris hypothesis of corporate takeovers. *The Journal of Business, 59*(2), 197–216.
Shoucri, M. (1991). Patriarchal election. *Azis Atiya The Coptic Encyclopedia* (pp. 1911–1912). New York: Macmillan.
Sintomer, Y. (2014). *Petite histoire de l'expérimentation démocratique: Tirage au sort et politique d'Athènes à nos jours.* Paris, France: La Découverte.
Stolz, P. (1986). Parteienwettbewerb, politisches Kartell und Tausch zwischen sozioökonomischen Gruppen. *Swiss Journal of Economics and Statistics (SJES), 122*(IV), 657–675.
Sturm, R., & Antoniakis, J. (2014). Interpersonal power: A review, critique, and research agenda. *Journal of Management, 41*(1), 136–163.
Van Reybrouck, D. (2016). *Against elections: The case for democracy.* London: Penguin Random House.
Weber, R., Camerer, C., Rottenstreich, Y., & Knez, M. (2001). The illusion of leadership: Misattribution of cause in coordination games. *Organization Science, 12*(5), 582–598.
Winnington-Ingram, R. P. (1971). The Second Stasimon of the Oedipus Tyrannus. *The Journal of Hellenic Studies, 91*, 119–135.
Woodruff, P. (2005). The shape of freedom: Democratic leadership in the ancient world. In J. B. Ciulla, T. L. Price, & S. E. Murphy (Eds.). *The quest for moral leaders: Essays on leadership ethics.* Northampton: Eldward Elgar Publishing.
Zehnder, C., Herz, H., & Bonardi, J.-P. (2017). A productive clash of cultures: Injecting economics into leadership research. *The Leadership Quarterly, 28*, 65–85.
Zeitoun, H., Osterloh, M., & Frey, B. S. (2014). Learning from ancient Athens: Demarchy and corporate governance. *The Academy of Management Perspectives, 28*(1), 1–14.
Zhang, H., Ou, A. Y., Tsui, A. S., & Wang, H. (2017). CEO humility, narcissism and firm innovation: A paradox perspective on CEO traits. *The Leadership Quarterly, 28*, 585–604.
Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics, 13*(1), 75–98.