

**WHO TO OPPOSE IF YOU SEEK SUPPORT:
ANTECEDENTS OF NEGATIVE PEER EVALUATIONS ON WIKIPEDIA**

HELGE KLAPPER

Rotterdam School of Management

HENNING PIEZUNKA

INSEAD

LINUS DAHLANDER

ESMT Berlin

ABSTRACT: When members of the same organization or community evaluate each other, we talk about peer evaluations. These evaluations occur continuously, meaning that that ego may evaluate a candidate today, and will themselves be evaluated tomorrow by their peers. While one could expect that ego – when expecting to be up for evaluation themselves – would avoid evaluating candidates *negatively* for fear of retaliation. We hypothesize that ego in fact chooses to evaluate *more* candidates negatively. We suggest that ego chooses to evaluate more candidates negatively despite the expected costs of retaliation because it allows ego to signal high standards and to uphold a balance of positive and negative evaluations that is accepted by their peers. Such negative evaluations, we argue, are directed toward candidates that minimize the risk of upsetting potential peers. We employ a unique dataset containing the evaluations of each election for administrators on Wikipedia from 2003 to 2014. We find that if ego is due to be evaluated themselves in the near future, they adjust their current evaluations. Specifically, they are more likely to evaluate candidates negatively but refrain from opposing similar and highly engaged candidates and avoid elections in which the outcome is uncertain.

ACKNOWLEDGMENTS: We are grateful to Gianluca Carnabuci, Javier Gimeno, Joe Labianca, Matteo Prato, David Reetz, Brian Silverman, and seminar participants at Aarhus University, Academy of Management, Rotterdam School of Management, Imperial College London, St. Gallen and Technical University Hamburg-Harburg for comments on earlier drafts. All authors contributed equally.

INTRODUCTION

Peer evaluations are ubiquitous and have been studied in a range of settings, such as decisions over promotions in organizations (Ertug & Castellucci, 2014; Lamont, 2012), the granting of authority in communities (O'Mahony & Ferraro, 2007), or the appointment of awards in creative industries (Aadland, Cattani, & Ferriani, 2018a; Cattani, Ferriani, & Allison, 2014; Shymko & Roulet, 2017). The intent behind peer evaluations is to access and aggregate information about a candidate in order to come to a more comprehensive and objective evaluation. However, research suggests that evaluations may not simply capture peers' honest opinions about a candidate, as peers evaluating a candidate may be strategic in their evaluations. For instance, Aadland, Cattani, and Ferriani (2018b) show how peers evaluating one another engage in a gift exchange.

We focus on peers who are up for evaluation themselves soon. A forward-looking evaluating peer (hereafter: ego) may therefore evaluate a candidate positively hoping that the candidate will engage in reciprocal behavior and evaluate ego positively when ego is up for evaluation. Prior research on peer evaluations has built on the concept of reciprocity, helping to explain why ego increases the number of positive evaluations (Aadland et al., 2018b; Taggar & Brown, 2006). Correspondingly, one might expect that ego is less likely to evaluate candidates negatively in order to avoid negative reciprocity. In this paper we point to a complementary explanation to such a reciprocal argument, proposing that under certain circumstances for strategic reasons ego chooses to evaluate more candidates *negatively*.

To develop our theory on strategic negative peer evaluations, we build on research on social networks, specifically the insight that ties serve as prisms affecting how actors are seen by others (Gulati & Higgins, 2003; Podolny, 2001; Stuart, Hoang, & Hybels, 1999). We conceptualize evaluations by ego as ties that affect how ego is perceived by the broader audience of peers. We

thus extend the focus beyond the dyad comprised of ego and the focal candidate. We suggest that when ego evaluates a candidate, ego anticipates how that evaluation will affect how the broader audience of peers will evaluate them. We argue that a negative evaluation can positively affect how ego is seen by the broader audience of peers. Specifically, we suggest that negative evaluations help to signal high standards and to appear non-strategic by upholding a balance of positive and negative ties that is accepted by their peers. We thus expect that ego deploys negative evaluations strategically in order to manage how he is perceived the broader audience of peers (Bolino, Kacmar, Turnley, & Gilstrap, 2008; Bozeman & Kacmar, 1997; Higgins, Judge, & Ferris, 2003). We also suggest ego will be strategic about which candidates they evaluate negatively. Specifically, we suggest that ego will avoid evaluating candidates negatively where a negative evaluation could upset the broader audience of peers.

Testing our predictions requires detailed longitudinal information about both ego and the candidates, as well as about the evaluations. Such data, in particular data on negative peer evaluations, is almost impossible to find as it is highly sensitive. To overcome this hurdle, we use a unique data source from Wikipedia. Beyond the features that render Wikipedia a generally attractive research setting (Gallus, 2017; Klapper & Reitzig, 2018; Leskovec, Huttenlocher, & Kleinberg, 2010; Nagaraj, 2018; Piskorski & Gorbatai, 2017; Zhang & Zhu, 2011), we leverage that Wikipedia archives members' evaluations of peers who are candidates for an administrator position. We combine this information on the evaluations with rich time-stamped information on ego and the candidates. Most notably, we match this data with detailed information about all contributions made to Wikipedia. We analyze 871,644 evaluations made by 8,759 distinct egos on 2,870 candidates across 3,447 candidacies between 2003 and 2014. In sum, we have detailed

information about both ego and candidates, as well as information on who has evaluated whom and how, allowing us to shed new light on how negative ties can arise from strategic considerations.

In line with our theory, we find that when an ego plans to go up for evaluation, it *increases* the tendency to evaluate other candidates negatively. However, ego does so *selectively*. This means, they will only evaluate certain candidates negatively, assuming that the majority of peers will not be upset by negative evaluations of these candidates. Specifically, we find that ego avoids negatively evaluating candidates in cases where (a) the outcome of the election is uncertain, (b) the candidate is highly visible, and (c) ego and the candidate work in the same domain. A unique feature of our work is that we can connect these behaviors to outcomes and analyze the *consequences* of ego's evaluation of candidates. Specifically, we examine to what degree the strategic negative evaluations we theorize about and identify in our data actually help ego be more positively evaluated by their peers when they are up for evaluation themselves.

Our paper contributes to two streams of literature. The first contribution stems from integrating negative evaluations into studies of *peer evaluation*. Prior work has pointed out how strategic reasons lead actors to evaluate candidates *positively* in order to foster a relationship with some peers (Aadland et al., 2018a). Yet some peer evaluations are *negative*, and it has to date been unclear how such ties arise. In fact, prior research building on a reciprocal logic suggests that ego should avoid evaluating peers negatively. By contrast, we provide evidence that ego does evaluate peers negatively for strategic reasons. Our theory also helps to shed light on the frequently observed negative evaluations of distant peers (Boudreau, Guinan, Lakhani, & Riedl, 2016; Trapido, 2013): our study suggests that ego must provide negative evaluations to be positively perceived by the broad audience of peers; at the same time, ego is expected to have a positive relationship with peers in their domain. Our findings suggest that negatively evaluating distant

peers is an easy way to indicate high standards and to appear non-strategic even if ego is overly positive toward close peers they depend upon.

The second contribution is to *social networks*. A rich body of work has examined how actors navigate social networks in order to advance (Gargiulo, 1993; Kleinbaum, 2012; Krackhardt, 1990). We focus on how egos use negative evaluations strategically to shape how they are perceived by peers. Prior work on how social ties shape how actors are perceived has focused on ties that are *positive* and *mutual*, and serve as an indicator of an ego's quality and trustworthiness (Gulati & Gargiulo, 1999; Podolny, 2001). By contrast, the evaluations we focus on are *negative* and *one-directional* (i.e., the candidate does not have to agree to be evaluated), and inform peers about egos' standards. While the negative evaluations also serve the purpose of shaping how ego is perceived, they are different insofar as ego has more agency as such ties are outgoing and one-directional and as the underlying signal is different. Our study also highlights a so-far neglected trade-off actors face when managing their social network social network. Ego needs to manage not only direct relationships, but also how they are perceived by the broader audience of peers. While negative (positive) evaluations may harm (help) an ego's *relationship* with a focal candidate, they improve (hinder) how they are perceived by the broader audience of peers. Our study illustrates circumstances under which ego perceives it be beneficial to risk retaliation to improve their standing among peers. We also contribute to the important, but still understudied domain of negative ties (Labianca & Brass, 2006). In contrast to prior work that has often focused on the target of negative ties (Ellwardt, Labianca, & Wittek, 2012), we examine the motivation for the originator of negative ties. We illustrate how they choose to form negative ties for strategic reasons.

THEORETICAL BACKGROUND AND HYPOTHESES

Evaluations. An evaluation is an inherently social process surrounded by uncertainty (Lamont, 2012). A challenge that often arises is that a true objective value is difficult to discern. For instance, the qualifications of a person applying for asylum in the US may be difficult to evaluate (Chen, Moskowitz, & Shue, 2016), the value of innovation projects are hard to discern before hitting the market (Criscuolo, Dahlander, Grohsjean, & Salter, 2017), scientific grants can be difficult to allocate in practice (Boudreau et al., 2016), and pitches in baseball happen so quickly that it is difficult for an umpire to decide whether it is within the strike zone or not (Green & Daniels, 2014; Kim & King, 2014). In these uncertain situations, research has documented how evaluations can be biased as people fall prey to factors other than its pure quality (Boudreau et al., 2016; Criscuolo et al., 2017). For instance, in an experiment on unknown music bands, evaluators look at the opinions of others to make a judgment rather than only the song itself (Salganik, Dodds, & Watts, 2006). Even if evaluators want to be “fair judges and experts whose opinions matter” (Beljean, Chong, & Lamont, 2016: 42), the key insight from this literature for our work is that aspects other than the pure quality of the candidate affect evaluations. Evaluators follow established norms about what the audience would expect in a given situation (Chong, 2013; Lamont, 2012). In evaluating others, we may thus be largely affected by thinking of how others perceive us because of evaluations. This literature disregards who votes for whom but is more interested in the aggregate outcomes of evaluations. This is unfortunate, as the aggregate outcomes can thus be a result of different people’s strategic consideration rather than the merit of the candidate.

Peer evaluations. Peer evaluations are inherently social with directed ties being formed between ego and candidates. These evaluations among peers are omnipresent in situations such as academia (Boudreau et al., 2016), the Academy Awards (Rossman, Esparza, & Bonacich, 2010),

music awards, and the advertisement industry (Aadland et al., 2018b), as well as in consulting companies, law firms, and organizations characterized by flat hierarchies (Puranam, Alexy, & Reitzig, 2014). A particular feature of these assessments is that one evaluates one's peers. Thus, it would be typical for ego to occupy the role of the evaluator as well as the role of the evaluated, that is, the same actor that evaluates peers may also be evaluated by their peers in the future.

Our research builds on the idea that ego's evaluation of peers will affect how ego themselves will be evaluated by their peers. Anecdotal qualitative data from our research context underscores this claim: we found multiple instances where peers evaluating a candidate explicitly commented on how the candidate had evaluated other candidates.¹ Ego may thus anticipate how the evaluations of a peer will affect how they themselves will be evaluated – and adjust their evaluations of candidates correspondingly. Being up for evaluation may thus affect (a) how many candidates ego evaluates, and (b) how ego evaluates the candidate (i.e., positively or negatively).

Positive Evaluations. Prior research points to one strategy that ego pursues when they themselves are evaluated in the near future: ego is likely to evaluate more candidates and to do so positively. Such positive evaluations allow ego to form – or to foster pre-existing – relationships. Building on the idea of positive reciprocity (Flynn, 2006; Gouldner, 1960; Taggar & Brown, 2006), ego hopes that candidates who they evaluate positively will evaluate them positively when they themselves declare their candidacy. Indeed, prior research shows that the pattern underlying peer evaluations resembles a gift exchange (Aadland et al., 2018b). This strategy where actors positively evaluate candidates in order to secure positive evaluations also underlies the inflation of positive

¹ For example, in the case of the promotion of user *Alkivar*, peers who engaged as part of the evaluation pointed out that the user had only given positive evaluations prior to their own evaluation in order to “butter up” votes. See https://en.wikipedia.org/wiki/Wikipedia:Requests_for_adminship/Alkivar_3.

evaluations that can be observed on transaction platforms such as Uber and Upwork, etc. (Filippas, Horton, & Golden, 2018). In line with this reciprocal logic, we do expect that when an ego plans to go up for evaluation, they are likely to evaluate more candidates and to do so positively, that is, they increase the number of positive evaluations.

Null Hypothesis. Building on the same reciprocal logic one may expect that ego also reduces the number of negative evaluations. If ego evaluates a candidate negatively, it may result in negative reciprocal behavior where the candidate retaliates and negatively evaluates ego when ego is up for evaluation (Fernandez-Mateo & Coh, 2015; Taylor, 1991). To be promoted, ego must not only secure positive evaluations, but must also avoid receiving negative evaluations; in fact, avoiding negative evaluations is particularly important given the high impact that negative evaluations have. As Labianca and Brass (2006) illustrate, the effect of positive and negative ties is not symmetric; negative ties are particularly impactful. For instance, a single negative tenure letter, a weak recommendation when applying for jobs or, as in our case, a single negative vote on Wikipedia, weighs heavier than a positive evaluation. Additionally, research on peer assessment in schools finds that without anonymity students are less likely to give critical feedback, indicating that they know the social costs of negative evaluations (Howard, Barrett, & Frick, 2010; Vanderhoven, Raes, Montrieux, Rotsaert, & Schellens, 2015). Building on a reciprocal logic one may thus expect that egos not only increase the number of candidates they evaluate positively, but they also decrease the number of candidates they evaluate negatively. We do, however, frame against this null hypothesis. We suggest that egos – when they go up for evaluation – will increase the number of positive evaluations, but, instead of reducing the number of negative evaluations, will instead increase the number of negative evaluations.

At the core of our argument is that the effect of social evaluations goes beyond creating a network of relationships that build on positive or negative reciprocity. They also have a different effect. Prior research on social network ties illustrates that ties serve as pipes and prisms, that is, they affect an actor's access to resources and shape an actor's external identity (i.e., how the actor is perceived by others) (Podolny, 1993; Podolny, 2001). The evaluations that ego provides are like pipes in the sense that they allow for the exchange of (dis)avors with the candidates that ego evaluates. According to this logic, ego should indeed increase the number of positive evaluations and decrease the number of negative evaluations. However, evaluations also serve as prisms in that they co-determine an ego's external identity, that is, how they are perceived by the broader audience of peers, as they are likely to pay attention to how ego evaluates candidates and will form their judgment based on that.

One way in which negative evaluations may benefit how ego is perceived is that such negative evaluations indicate that ego has high standards. If the broader audience of peers believes ego to have high standards, it is more willing to support them because peers with high standards are needed to uphold the quality of the overall community. Having high standards and being critical of others also indicates that ego is skilled and capable because ego recognizes the flaws in a candidate's work that are not visible to everyone (Amabile, 1983).

Another reason why ego evaluates more candidates negatively when they plan to go up for evaluation themselves are that peers could perceive positive evaluations of candidates as a way to foster reciprocity (Gagne, Khan, Lydon, & To, 2008), and may correspondingly perceive an ego as political, unethical, and strategic (Blank, 2006). Interestingly, and somewhat paradoxically, egos may thus *strategically* evaluate candidates negatively so as to be seen as *non-strategic*. It is likely that the audience will not take into account each evaluation individually but rather the portfolio of

evaluations (Bowers, 2015). If ego evaluated everyone positively it would give the audience little to determine ego's profile and preferences, that is, the signal of the evaluation would be very low. By balancing the amount of positive and negative evaluations, ego can show that their evaluations are credible (Bowers, 2019).

We thus theorize that the null hypothesis that ego avoids negative evaluation does *not* hold. In contrast to this argument, we hypothesize that ego is in fact *more* likely to evaluate candidates negatively. We suggest that ego chooses to give these negative evaluations and anticipates that negatively evaluated candidates are likely to reciprocate and evaluate ego negatively when they are up for evaluation. Ego is willing to pay the price of invoking negative reciprocity because ego's focus is not on managing the relationship with the focal candidate, but on managing how they are perceived by the broader audience of peers.

Hypothesis 1 (H1): *When ego plans to go up for evaluation themselves, the tendency to evaluate a candidate (a) positively as well as (b) negatively increases.*

In H1b we suggest that ego will evaluate more candidates negatively when they are up for evaluation. This assumes that ego expects that negative evaluations improve how ego is perceived by their peers, and that the corresponding benefits outweigh the costs of the likely negative reciprocity. This hypothesis is subject to an important boundary condition. The support that ego can mobilize for their promotion from the broader audience of peers by evaluating a candidate negatively must outweigh the opposition that ego is likely to face from the negatively evaluated candidate.

While our argument rests on the idea that negatively evaluating candidates can help to improve how ego is perceived, we also believe that negative evaluations of candidates come with the risk

of offending the broader audience of peers; peers may disagree with ego's negative evaluation of a candidate. If so, ego's negative evaluation would substantially worsen their chances when they are up for promotion, because they would be likely to experience negative reciprocity from the negatively evaluated candidates and face opposition from an offended broader audience if peers. We therefore hypothesize that ego – while generally increasing the number of candidates that they evaluate negatively (H1b) – does so *selectively*. In what follows, we point to three instances where we suggest that ego forgoes evaluating a candidate negatively to avoid offending peers.

Avoiding negative evaluation of candidates where outcome is uncertain. We theorize that ego only negatively evaluates candidates where the overall outcome of the candidate's evaluation is relatively certain and it is not difficult for ego to predict how other evaluators assess candidates (Correll, Ridgeway, Zuckerman, Jank, Jordan-Bloch, & Nakagawa, 2017; Ridgeway & Correll, 2006). In an evaluation where the outcome is uncertain, a negative evaluation by ego is more likely to be pivotal, that is, ego's negative evaluation may result in the rejection of the candidate.² For peers who support the focal candidate, ego's negative evaluation is particularly bothersome. While peers may appreciate ego for being critical of candidates for the reasons outlined in H1, they probably do not want ego to be critical of candidates they appreciate. They are likely to be aware and to remember a negative evaluation that (almost) resulted in the rejection of a candidate they supported – and would be more likely to hold that against ego when ego herself is up for evaluation. However, this case is not necessarily true for the reverse—evaluating a candidate positively who

² Negative evaluations tend to have a higher impact than positive evaluations in general because organizations often require a supermajority (i.e., the threshold is above 50%) so that the marginal effect of negative evaluation is higher than the marginal effect of positive evaluation.

is then promoted—since negative evaluations are likely to evoke stronger responses in the audience (Labianca & Brass, 2006; Taylor, 1991). We suggest that ego is aware of this risk that comes with playing a pivotal role.³ Research on blame avoidance (Weaver, 1986) shows that actors minimize engaging on issues where they could be blamed for the outcome (Balla, Lawrence, Maltzman, & Sigelman, 2002; Vis & van Kersbergen, 2016). We thus suggest that ego will avoid evaluating candidates negatively if the outcome is uncertain.

***Hypothesis 2 (H2):** While ego is more likely to negatively evaluate peers when planning to go up for evaluation themselves (as hypothesized in H1b), ego is likely to avoid negatively evaluating candidates where the outcome of the evaluation is uncertain.*

Avoiding negative evaluations of candidates who are very active. We hypothesize that one way in which ego minimizes the risk of offending peers is *to avoid the evaluations of candidates who are very active* (i.e., they have contributed extensively). A very active candidate is likely to be appreciated by many peers and to have high status among peers. If ego negatively evaluates such a candidate, she is more likely to offend peers. Peers may take ego's negative evaluation of a highly active candidate as an indicator that ego does not appreciate the activities of others or is unable to recognize the value of the candidate. Peers may also want to retaliate against ego to signal their deference/allegiance to the highly active candidate. Moreover, ego may also be aiming to raise the bar, that is, to increase the standards that are used to evaluate candidates. Prior research

³ In such an uncertain evaluation, no matter whether ego evaluates the candidate positively or negatively - it will always result in dissent with a substantial share of their peers. A rich body of research has pointed to the challenges that evaluators have when they are confronted with contradicting expectations from a not-aligned audience with heterogeneous expectations (Meyer & Rowan 1977). They may thus avoid participating in such an evaluation altogether. If, by contrast, the evaluating peers were uniform in their evaluation of the candidate, it is easy for ego to avoid dissent by simply evaluating the candidate in line with how others evaluate the candidate.

has shown that one goal actors pursue is to use their evaluations to create or to set standards (Bourdieu, 1993). If ego negatively evaluates a highly active candidate, peers may use the same high standards against ego, when they go up for evaluation. To avoid offending numerous peers and to increase the standards against which they are measured, we suggest ego avoids negative evaluations of candidates who are highly engaged.

Hypothesis 3 (H3): *While ego is more likely to negatively evaluate peers when planning to go up for evaluation themselves (as hypothesized in H1b), they are likely to avoid negatively evaluating candidates who are highly engaged.*

Avoiding negative evaluations of candidates who work in similar domain. We hypothesize that one way in which ego minimizes the risk of offending peers is *to avoid the negative evaluations of candidates who work in the same domain*. At first sight, one may read prior research in a way that would suggest that ego has a higher tendency to evaluate those candidates negatively that work in the same domain as ego. Studying book reviews in American newspapers, Chong (2013, 2015) finds that evaluators try to be critical of similar candidates who are similar to ego in order to appear impartial and make legitimate judgments. Similarly, Boudreau et al. (2016) find that grants evaluators are stricter toward proposals that are closer to their expertise, and that “specialized experts provide more discerning evaluations but they also provide systematically lower – and more critical – evaluations.” However, the setting of these studies differs from ours insofar as in our setting ego and the candidates are supposed to collaborate to foster “their” domain. If, in a peer-based evaluation within an organization, ego does not appreciate their co-worker, it may in fact reflect negatively on ego. When ego negatively evaluates candidates who work in the same domain as ego, peers may interpret this as an indicator that ego has failed to establish good relationships

within the community. They may even conclude that ego wants to undermine a certain candidates' promotions to gain momentum in that domain. Others might perceive this as a "turf fight" in which ego is trying to defend their power (Zuckerman, 2004), and such turf wars are unlikely to be appreciated.

***Hypothesis 4 (H4):** While ego is more likely to negatively evaluate peers when planning to go up for evaluation themselves (as hypothesized in H1b), they are likely to avoid negatively evaluating candidates who work in the same domain.*

METHODS

Research context

We test our theory on peer evaluations in the context of Wikipedia (Gallus, 2017; Klapper & Reitzig, 2018; Leskovec et al., 2010; Nagaraj, 2018; Piskorski & Gorbatai, 2017; Zhang & Zhu, 2011), the largest free online encyclopedia. Since 2007, Wikipedia has been in the top 10 of the most visited websites worldwide. An important feature of Wikipedia is that all users, whether registered or unregistered, can edit pages and contribute knowledge to the various articles. As a result of this broad participation, and to coordinate user activity, Wikipedia has implemented several coordination mechanisms: most prominently making certain users administrators. Administrators enjoy additional privileges such as the right to (un)block user accounts, restrict page edits, and delete pages.

We examine the elections in which members are elected to be administrators. Any registered user, who we call member, can apply, or "request," to become an administrator. This election, officially called "Request for Adminship" (RfA), allows any unregistered user to participate by asking questions and lets registered members evaluate candidates by voting on the request.

An election can be initiated by a community member nominating another community member as a candidate, or through a self-nomination. Candidates are then asked to answer questions about their past activities, plans, and opinions about Wikipedia's processes. All members are allowed to submit questions. Potential evaluators can inform themselves about the candidate by checking their user page, akin to a private webpage on Wikipedia, and past contributions, which are all transparent for anybody to view. For example, they can trace the full voting record of the upcoming candidate. All members are then allowed to evaluate the candidate, which can take the form of supporting, opposing, or staying neutral regarding the request.

The nominal vote is, however, only a part of the overall vote. Members also justify their evaluation. For example, members negatively evaluate a candidate due to lack of experience, controversial edits or discussion entries. This can lead to further discussions, since other members are allowed to respond to an individual's evaluation. Negative evaluations, in particular, are often starting points for further discussions. This enables us to qualitatively compare the reasons cited by members with the reasons uncovered by the quantitative analysis.

Elections usually run for seven days, unless the candidate withdraws the nomination early. The number of evaluators participating in an election is anywhere between 1 and 388, although there is no formal upper limit. Officially, bureaucrats – a small group of elected Wikipedia members with even further privileges – determine whether there is consensus regarding the nomination. Typically, an approval rating of more than 75 percent will lead to a successful election, and below 65 percent will not. Within the range of 65 to 75 percent, bureaucrats use their discretion to determine whether there is consensus among the group. This implies that negative evaluations are *very* important in affecting the final outcome of the election, because already 25 percent of the evaluations against a candidate trigger the intervention of bureaucrats. Another important feature

of the election is that there is no limit to the number of members who can be promoted to become administrators, therefore candidates are not running against each other.

The process for becoming an administrator is not simple. Whereas it was originally considered by Wikipedia's founder as merely a technical support function ("not a big deal"), the importance of the position and the scrutiny of evaluations for promotions has increased (Jemielniak, 2014; Kittur, Suh, Pendleton, & Chi, 2007). Over time, candidates have been asked to provide detailed information on their activity and answer questions from anyone who is interested. In order to make the process more vivid, consider these two examples. The first example of a request in 2007⁴ shows an experienced candidate, "*Cuchullain*," who answers a few questions in the beginning and receives support from 60 evaluators without encountering any real opposition. But requests for becoming an administrator can take a more controversial turn, as can be seen in the second example.⁵ In this case, when the user "*Alkivar*" requested adminship for the third time after being rejected two times they had to answer seven questions about the potential duties and activities as an admin and about their past activity. At the beginning, a majority of members initially supported the request. However, considerable opposition formed. Members opposed the candidate for not doing the work properly, for example, not summarizing edits properly or being unfriendly to other editors. A member accused *Alkivar* of supporting more recent candidates, while having opposing more candidates in the past, in order to "butter up" support. In the end, the request was approved by a narrow three-fourths majority, not requiring a further decision by bureaucrats.

⁴ https://en.wikipedia.org/wiki/Wikipedia:Requests_for_adminship/Cuchullain

⁵ https://en.wikipedia.org/wiki/Wikipedia:Requests_for_adminship/Alkivar_3

An important feature of the election is complete transparency: any member can see all the comments and votes of potential candidates. Even though most members do not use their real identity, the pseudonym represents their Wikipedia identity and cannot be easily changed. We thus use all the data from the elections where we have full information about each candidate and the type of votes cast by each member. We combine this data with other information from Wikipedia on the pages each individual has edited, the overall activity on those pages, and previous votes by each individual.

Figure 1 shows the number of elections over time and the average number of members participating in an election. The number of requests peaked in 2007 and then dropped off significantly. This is likely due to the increased scrutiny in the election process and a saturation of the number of required admins (Jemielniak, 2014). The average number of members participating in an election increased from 10 to about 60. Considering there are more than 10,000 active members, this is a rather low number, emphasizing the importance of not only investigating the types of evaluations but whether a member votes at all, which our empirical strategy accounts for.

Insert Figure 1 about here

Our final sample consists of 3,447 elections with 871,644 potential evaluations made by 8,759 members on 2,870 candidates. About 46 percent of the elections are successful, indicating that the evaluations of members are of great importance. The first observable election took place in 2003. We track elections from 2003 until the end of 2014.

Data analysis

In our analysis, we focus on the decision of the focal member (i.e., ego) whether and how to evaluate a candidate as the unit of analysis. Members have the freedom but not the obligation to

evaluate any candidate. In order to analyze whether and how a Wikipedia member evaluates a candidate, we compare the evaluating members to non-evaluating ones. We create a risk set of members, which comprises all members who could have participated in the focal evaluation. For this we include all members who had been active in the time before the election, that is, had edited at least one Wikipedia page, and who at one point during their tenure had shown an interest in evaluating candidates operationalized by including only those members who had participated in the evaluation of another candidate. For our analysis, for each focal election we select all evaluating members, and for each evaluating member we include five members who did not evaluate that candidate but were part of our risk set and already active before the given election. This choice-based sampling strategy has been used extensively in the studies of network formation (e.g. Hallen, 2008; Sorenson & Stuart, 2001). When we came across an election with such high participation that there were not enough non-participating members to fill the required risk set “five-unobserved-votes-per-observed-vote,” all available non-participating members were included. Such cases were very rare in the data as illustrated by the average of 4.8 non-participating members for one participating member in our data.

Variables

To construct our variables, we collect information about the ego, the candidate, the election, and the dyad of ego and candidate. Unless otherwise specified, all variables are constructed by month. If a variable considers past activity, we generally “look back” six months prior to an evaluation. As explained in detail in the Appendix, this is motivated by both substantive and empirical considerations.

Dependent variable

Evaluation. Our dependent variable captures two aspects about ego’s evaluation of a candidate.⁶ First, we determine *whether* ego participated in the evaluation of the candidate, using the binary variable *Participation*. All registered users on Wikipedia can self-select to evaluate every candidate who is up for evaluation. Second, if ego evaluated the candidate, we can classify the participation as: (1) positive or (2) negative.⁷

Independent variables

Up for evaluation. We measure whether ego plans to go up for evaluation by using a dummy (*Up for evaluation*) that takes the value 1 for all evaluations in which ego could potentially participate six months before their election. We assume that in this time frame ego anticipates that their actions will be scrutinized when they go up for election. We verify whether the six-month window captures any or all anticipation in two ways. First, we qualitatively checked comments made during the elections by evaluators. Although no specific time window was established, evaluators mentioned that three or four months were not enough to fully evaluate the activity of a candidate. Second, we investigate empirically for how long the anticipation by ego can be traced (see Technical Appendix T1). The results indicate that potential candidates increased their evaluation activity at least six months before an election.

We thus expect ego to be particularly conscious about their actions in this timeframe. Even though ego does not have to state in advance whether they plan to declare their candidacy for becoming an admin, active members are aware of the complicated and difficult process of

⁶ Whereas we use the second evaluation variable for our main analysis, we use Participation for our robustness checks.

⁷ Members also have the option to vote “neutral.” In our main analysis we exclude this category for clarity. All results hold when that category is included.

becoming an admin (Jemielniak, 2014). Therefore, we assume most evaluators plan ahead to prepare for requesting a promotion.

Moderator variables

Uncertain outcome (H2). We count elections in which the share of positive evaluations is between 65 percent and 75 percent as having an uncertain outcome; if the share of positive evaluations is above 75 percent or below 65 percent the candidate is promoted, respectively rejected, without further discussion. In the case of 65 percent to 75 percent the decision goes to bureaucrats, which are members of the Wikipedia leadership team with special decision rights, who decide whether the candidate is promoted. Thus, if the current share of positive evaluations is between 65 percent and 75 percent the outcome is even more uncertain than in other elections, since it depends on an additional decision outside of the promotion request. In such a case, it is difficult or almost impossible for ego not to offend peers. We measure a close election with a dummy taking the value 1 if the approval rate evaluation at the end of the evaluation period is between 65 percent and 75 percent.⁸

Candidate activity (H3). We measure the activity of the candidate by calculating the share of total page edits on Wikipedia in the lookback period of six months. Because the share of the candidate's page edits in relation to the edits by all members is very low given Wikipedia's size, we standardize this variable.

Domain similarity (H4). The measure captures the degree to which ego and the candidate have edited the same pages. To compute this measure, we construct a vector for the potential evaluator

⁸ Because we do not know when non-evaluations were made, we cannot employ a time-variant measure of outcome uncertainty. As a robustness check, we use "average" outcome uncertainty, calculating what value the approval rate took on average over the course of the election. The results remain robust.

and the candidate that includes *all* Wikipedia pages. In these vectors, 1s indicates whether a page was revised by the respective actor during the lookback period of six months, and 0s indicates that a page was not revised by them during the lookback period (Figure 2). We then calculate the correlation coefficient for the vectors of the ego and the candidate.

Insert Figure 2 about here

We include control variables that may affect the tendency to, and how to, participate in a vote. The variables can be categorized into the level of (1) the ego, (2) the candidate, and (3) their dyadic relationship, which we elaborate on in turn below.

Control variables: Ego

Members differ in their tendency to participate in evaluations. Some members are more active than others, independent of the particular candidate or promotion decision. ***Ego's tendency to participate in the evaluation of a candidate*** calculates the number of elections in which the ego has registered a valid, non-withdrawn evaluation divided by the total number of elections in which ego could have participated in the lookback period. The overall amount and type activity of ego might also affect the tendency to evaluate the focal candidate. Specifically, we control for ego's experience in evaluating candidates by measuring the ***Number of evaluations they participated in*** before the evaluation of the focal candidate. ***Ego's edited pages*** is the logarithm of the number of unique pages that ego revised during the lookback period. Next, we measure the activity of a potential in relation to the overall activity on Wikipedia. ***Ego's engagement*** is their share of total page edits in the lookback period. This variable indicates how active ego was in comparison to other members. Because the share of their page edits in relation to the edits by all Wikipedia members is very low, we standardize this variable. We also control how prominent ego is on the

Wikipedia pages they edit. A member might be more active on pages with few other members, and thereby less likely to miss the work of candidates also active on those pages. The variable *Share of edits* indicates how active ego was on edited pages compared to other editing members. For example, a member could be very active on many pages by contributing little on each or could edit a lot on very few pages. To calculate the variable, we use the pages ego edited in the lookback period and take the mean of the share of edits. For each page that ego edited in the lookback period, we calculate the share of edits divided by total edits. We then calculate an average of these values. We also control for *Tenure* calculated as months since ego's registration. These variables capture the time-varying characteristics of an individual due to engage in evaluations.

Ego can also be characterized in terms of their previous promotion decision. If ego has been successfully promoted, ego has become an administrator (*Admin*), which can impact their criteria and likelihood for voting. The other possibility is that they have tried to become an admin, but have failed in the past (*Previous failed candidacy*). In order to classify ego's evaluation pattern more, we calculate an ego's *Tendency to reverse evaluations* measured as how often they changed their positive evaluation to a negative one (or vice versa) during an election. We also calculate ego's likelihood of deviating from the majority position (*Deviation from majority*) and at what relative point in time they tend to participate in the evaluation vote (*Relative order to vote*) measured as the place in the voting chain. In measuring these variables, we account for differences between ego's general tendency to participate in evaluations, and for voting negatively.

Control variables: Candidate

We collect the corresponding information about the candidate, namely the *Candidate's edited pages*, *Share of edits*, *Previous failed candidacy* and *Tenure*. We also account for whether the candidate evaluated other candidates in the lookback period with a dummy *Prior evaluation*. In

the context of Wikipedia, two aspects matter to members during elections when evaluating candidates: the number of reverts, edits that delete previous changes, by candidates and the ratio of edits that have been commented. *Edits reverted* by the candidates is measured as the ratio of reverts to overall edits in the lookback window. Reverts can refer to cleaning up unproductive edits by previous users, for example, vandalism, or could mean an overzealous deleting of edits that the candidate did not like. A very high ratio of reverts can be seen as controversial for candidates, and would thus lead to opposition. Additionally, when users edit pages, they can comment on their edit to explain to future users what they have changed.⁹ This facilitates coordination among members. Members may use the ratio of edits that were commented (*Edits commented*) as an indicator of the thoroughness of candidates, which could positively affect their likelihood of receiving a vote.

Control variables: Dyad Ego-Candidate

We control for previous interactions in promotion decisions. Importantly, if the current candidate voted in the ego's prior election, that vote could have repercussions. For example, if the current candidate evaluated ego negatively in the past, the ego might be inclined to evaluate the focal candidate negatively in line with negative reciprocity. We control for this by using a dummy, *Prior neg. evaluation of ego by candidate*. This binary variable indicates whether the focal candidate voted in a past election where ego was the candidate and the focal candidate voted "oppose": 1 if such a vote took place, 0 otherwise. We also control for *Prior pos. evaluation of ego by candidate*, indicating whether the focal candidate voted in a past election where ego was the candidate and the focal candidate voted "support." Finally, we control for how often ego and the candidate have deleted the other's edits as an indicator of potential prior disagreements. *Edits*

⁹ This is similar to version notes or comments in programming code.

of candidate reverted takes the logarithm of edits of the candidate that were reverted by ego, whereas *Edits by candidate reverted* takes the logarithm of edits of ego that were reverted by the candidate.

Estimation strategy

We test our hypothesis by estimating a multinomial logistic regression. The unit of analysis is ego's decision to evaluate a candidate positively or negatively or not at all.¹⁰ A multinomial logistic regression captures the odds of selecting one option without assuming an ordinal ranking of the options. This way, a positive effect of an independent variable on the evaluation decision, for example, a positive one, does not necessarily imply a negative effect on an alternative decision, for example, a negative evaluation. The odds are always in comparison to a reference group, in our case to a non-evaluation. We cluster standard errors on each evaluator to account for possible non-independence as voters occur several times.

An additional challenge is that for non-evaluations is not clear whether the ego was not *aware* of the election, for example, due to inactivity versus actively deciding *not* to evaluate a candidate. This could bias our selection sample and the results if we cannot distinguish the two explanations. Consider the example that the activity of a candidate has a positive effect on ego's decision to positively evaluate her; it could be partially the effect that the activity of the candidate made ego aware that the election exists. In order to compare active decisions to participate in the evaluation and active decisions not to, we need to restrict the sample to only non-evaluations where ego knew about the election but chose not to participate. Because we cannot observe whether potential

¹⁰ In our main analysis we exclude the neutral vote category for clarity. All results hold when that category is included.

evaluators were aware of the election, we use propensity score matching (Rosenbaum & Rubin, 1984) to find a subset of evaluations in which ego could have participated, and evaluations in which ego actually participated. The idea is to construct a credible counterfactual by calculating the propensity for each potential evaluation to result in an actual evaluation. This way we can select non-evaluations that are similar to actual evaluations in their observable factors. We select factors that we assume are correlated with awareness of the election: namely the activity of the potential evaluator, the prominence of the candidate, and the similarity between evaluator and candidate (for a full list of the factors, see Appendix T4.1). We then use a logistic regression to determine how these factors explain whether potential votes would result in votes and use this to calculate the propensity for a vote (Appendix T4.2). We then match a maximum of one non-evaluation to each evaluation, that is, nearest neighbor matching ($k=1$) with replacement. In order to evaluate the effectiveness of the propensity score method, we compare the sample means of the matched non-votes to actual votes (Appendix T4.3) and compare the distribution of propensity scores before and after matching (Appendix T4.4).¹¹ With this method, we end up with a way in which we can rule out awareness, enabling us to identify the effect of the decision to evaluate. In the end, we arrive at 243,981 evaluations and comparable non-evaluations.

RESULTS

Table 1 provides descriptive statistics for elections as well as variables on ego, candidate, and the dyad level. On average, 50 evaluators participated in an election with 37 positive evaluations. In Table 2, we report the average marginal effects from the multinomial logistic regressions on the

¹¹ As a robustness check we also use coarse exact matching as an alternative matching method to PSM. Our results remain robust.

likelihood of opposing or supporting a candidate in contrast to not voting. Model 1 is the baseline model, Model 2–4 add the interaction variables, and Model 5 is the full model with all interactions presented jointly. Each multinomial regression reports two columns – the first one for opposition and the second for endorsements compared to the baseline. Before discussing the main results, some control variables give an indication of the evaluation process. For example, higher tenure and activity (edited pages) by ego lead to an increase in negative evaluations, showing a rise in the standards of more active and tenured members. Furthermore, active and more tenured candidates are seen as more positive, while due diligence (edits commented) is also seen as positive. Finally, on a dyadic level the effect of positive reciprocity and negative retaliation is strong. If a candidate had negatively (positively) evaluated ego in a prior evaluation, ego is likely to respond in kind.

Table 1 and Table 2 about here

We find evidence for Hypothesis 1, which argues that if ego is up for evaluation they are more likely to evaluate a candidate more positively (beta = 0.587, relative risk ratio (rrr) = 80%, $p < 0.001$) and negatively (beta = 0.643, rrr = 90%, $p < 0.001$) in comparison to not participating in the evaluation.¹² The marginal coefficients for negative and positive evaluations are positive and significant, the probability for negative evaluations changes from 10 percent to 13 percent ($p < 0.001$) and for positive ones from 36 percent to 48 percent ($p < 0.001$) (see Figure 3).

¹² Note that coefficients in multinomial logistic regression have only limited value. Due to multiple possible outcomes that are compared to the base outcome, a marginal increase does not necessarily require an overall increase in likelihood. For example, if the coefficient for “under evaluation” is much larger for a positive evaluation than a negative one (while both are positive) this could lead to an overall decrease in likelihood of negative evaluation. It is thus crucial to compare overall probabilities, which we show in Figure 3 to Figure 5.

Insert Figure 3 about here

We next test whether an ego who is up for evaluation reduces their negative evaluations in a close election. The marginal effects are significant ($\beta = -0.414$, $\text{rrr} = -34\%$, $p < 0.001$). Whereas a member under evaluation is as likely to negatively evaluate a candidate in uncertain elections, that is, there is no significant difference ($\text{diff.} = -0.001$, $p = 0.854$), they are far more likely to negatively evaluate candidates in “certain” elections ($\text{diff.} = 0.033$, $p < 0.001$). Figure 4 shows the probabilities due to scrutiny and illustrates that there is no difference between an ego who is up for evaluation and those who are not in situations where the outcome is uncertain. However, when a focal vote has less impact on the outcome, then an ego who is up for evaluation is more likely to oppose. This supports Hypothesis 2 by showing that evaluators strategically place their negative votes when their vote is unlikely to swing the election in either direction.

Insert Figure 4 about here

Hypothesis 3 posits that when ego is up for evaluation they are less likely to negatively evaluate active and engaged candidates. The marginal effects are significant ($\beta = -0.118$, $\text{rrr} = -11\%$, $p = 0.001$). Evaluators who are up for evaluation tend to oppose candidates that are not visible ($\text{diff.} = 0.034$, $p < 0.001$), but are marginally less likely to negatively evaluate engaged candidates when they are up for evaluation ($\text{diff.} = 0.019$, $p < 0.001$). In line with our expectations, Figure 5 shows the probabilities and illustrates a downward slope for when evaluators are under scrutiny and a candidate’s engagement increase in the likelihood of opposition. Specifically, Figure 5 indicates that actors under scrutiny are more likely to oppose non-visible candidates, whereas the difference between actors under scrutiny and non-scrutinized actors is smaller.

Insert Figure 5 about here

Finally, Hypothesis 4 proposes that when ego is up for evaluation they are less likely to negatively evaluate candidates from the same domain. The marginal effects are significant (beta = -0.063, rrr = -6%, p=0.045). An ego who is up for evaluation is marginally more likely to negatively evaluate candidates that are similar to other evaluators (diff. =0.042, p = 0.001), and they are far more likely to negatively evaluate distant candidates when up for evaluation (diff. =0.022, p < 0.001). Figure 5 shows the probabilities. With an election upcoming, as similarity increases between ego and the candidate, they are significantly less likely to oppose the candidate. Otherwise, however, this effect is much weaker.

Insert Figure 6 about here

Robustness checks and additional analysis

In the extensive Technical Appendix, we perform a series of robustness checks regarding (1) how long evaluators anticipate their own evaluation, (2) controlling for unobserved evaluator effects, (3) matching strategy, (4) main variable specification, and (5) control or comparison groups (see the Online/Technical Appendix). The reasons behind the robustness checks are explained in detail in the Appendices. In sum, our extensive robustness checks show that our results hold up for different specifications, alternative ways to specify the variables, and different time horizons, which lend additional support to our reasoning.

As an additional analysis we investigate how the timing of the evaluation for elections with an uncertain outcome is affected by being under evaluation. Our hypothesis (H2) argues that the uncertainty of the situation makes ego reluctant to evaluate the candidate negatively. By the same token, we can extend the argument that in such cases ego would wait longer with their evaluation

in order to reduce the perceived uncertainty of the outcome, since more evaluations by other members will have been made. Therefore, we regress the relative outcome certainty on the relative timing moderated by the type of evaluation (positive or negative) and whether ego is under evaluation (see Technical Appendix T5 for the regressions). Figure 7 shows the marginal effects of that regression. Importantly, when ego is under evaluation (right figure) they do an evaluation later than if they are not under evaluation if the outcome is uncertain. Thus, we find evidence that ego not only adjusts their propensity to evaluate candidates negatively but also adjusts the timing.

Insert Figure 7 about here

Post-hoc analysis: Linking hypothesized strategic behaviors to individual outcomes

In our theory of peer evaluation we hypothesized how ego evaluates a candidate. Our overarching argument was that ego evaluates candidates strategically to positively affect how ego themselves will be evaluated. A major strength of our study is that it allows us to analyze whether the behaviors we study actually result in the desired outcome of being positively evaluated themselves in the future. Therefore, as an additional step, we investigate whether the tactics used by evaluators under scrutiny pay off. Because we observe how the different egos under scrutiny vote and can determine whether they eventually become elected themselves, we can estimate the effectiveness of the behaviors we hypothesized and found. Table 3 shows an OLS regression with an approval rating measured as the number of support votes divided by the number of supporting and opposing votes as the dependent variable.¹³ We quantified each strategy, namely how often an ego under evaluation evaluated candidates positively (H1a) and negatively (H1b) in elections, the

¹³ We also run a logistic regression with success as a binary variable and find supporting results.

share of close elections in which they opposed (H2), the share of negative evaluations of engaged candidates (H3), and the share of negative evaluations of candidates from a similar domain (H4). Because not all evaluators under evaluation had the opportunity to evaluate candidates in elections with an uncertain outcome, or evaluate similar or visible candidates, the sample size is lower than the total number of members under evaluation (N=1,892). Technical Appendix T8 describes the construction of the variables in more detail. The results indicate that, on average, positive evaluations increase the approval of a request for adminship, whereas negative evaluations decrease it. However, when including a dummy to indicate whether a focal candidate at least once evaluated somebody negatively, the dummy has a positive effect. The results indicate that some negative evaluations are positive for being promoted but the marginal effect of more negative evaluations is negative. Moreover, negative evaluations in elections with an uncertain outcome and the opposition of similar or engaged candidates reduce the approval rating. Overall, the results indicate that the activities studied in H1, H2, H3, and H4 appear to be effective in becoming promoted.

Insert Table 3 about here

DISCUSSION

When seeking support – whom do you oppose? A dominant view consistent with prior work is that – in order to appear active and foster relationships with candidates – ego gives more evaluations that are positive. An alternate and much overlooked perspective is that negative ties and positive ties have to be considered in tandem. In line with our theoretical arguments – and more surprisingly – ego’s tendency to evaluate candidates *negatively* also increases when an evaluator seeks support. This finding departs from the generally accepted and mainly dominant

notion of reciprocity, which suggests that ego avoids evaluating candidates negatively. Instead, we found that ego provides *more* evaluations that are negative, albeit in a *selective* fashion. Negative evaluations are less likely to occur in situations where the general tendency is to provide evaluations that are more negative, where the outcome is uncertain, where the candidate is very engaged, and where the candidate works in a similar domain. Altogether, our work paves a new way for studying peer evaluations where the strategic considerations of the evaluator are paramount. These findings have broader theoretical implications, which we elaborate on below.

Theoretical implications

Peer evaluation. Prior work has pointed out how strategic reasons lead actors to evaluate candidates positively, namely in order to foster the relationship with some peers hoping that favors will be reciprocated (Aadland et al., 2018a). Despite the literature's one-sided focus on positive ties, it is an empirical fact that some peer evaluations are indeed negative. Why is that so? And this is where our first contribution comes in: to explain the antecedents of negative peer evaluations, we develop an argument that ego's tendency to evaluate candidates negatively can arise from strategic considerations. A key insight that emerges from our work is that ego does not evaluate candidates based on merit alone (Lee & Cole, 2003; Metiu & Kogut, 2001). In other words, each individual evaluation subject to strategic considerations is aggregated to an outcome that reflects their own ambition as much as the candidates' quality. An emerging body of work points out the important role of strategic considerations (Aadland et al., 2018a). We expand on prior work by showing that the deployed strategies are more complex as ego needs to cultivate relationships with candidates (and uses positive evaluations for that purpose). Furthermore, ego also needs to consider how the broader audience of peers perceives them (ego). For that, ego uses negative evaluations to manage perceptions of how they are viewed by their peers.

Our work sheds light on the frequently observed negative evaluations of distant peers (Boudreau et al., 2016; Trapido, 2013). A key insight is that ego must provide negative evaluations to indicate high standards. At the same time, ego is expected to maintain positive relationships with peers in their domain. Our findings suggest that negatively evaluating *distant* peers is an easier and less risky way to continue indicating high standards even if ego is overly positive toward close peers they depend upon. This signals that a person has set high standards, yet minimizes the risk of upsetting others. This can thus explain why distant peers often struggle to gain recognition. To use Merton's language, "marginal men" who are distant from a focal individual struggle to cut through the noise because their ideas are different, *and* are more likely to fall victim to people's strategic considerations.

Social networks and negative ties. A rich body of work has examined how ego's position in a network foster their own inter-organizational advancement (Burt, 2004; Casciaro, 1998; Gargiulo, Ertug, & Galunic, 2009; Jonczyk, Lee, Galunic, & Bensaou, 2016; Kleinbaum, 2012; Krackhardt, 1990; Srivastava, Goldberg, Manian, & Potts, 2018). Our study highlights an important and previously overlooked dilemma, which provides a new perspective. The actions that ego would need to take to foster the relationship with candidates are different from the actions that ego needs to take to improve how ego is perceived by the broad audience. Our study illustrates circumstances under which ego perceives the risk of retaliation as beneficial if it improves their standing among peers.

When their peers evaluate ego, one challenge ego faces is that they cannot control which peers will be doing the evaluation. Because peers self-select into evaluations by choosing whom to evaluate, the specific subset of peers that will evaluate them is *ex ante* unknown to ego and outside of ego's control. Ego thus needs to manage how the broad audience of peers view ego, as any one

member of the audience can carry out the evaluation on ego in the future. This suggests that despite the danger of negative reciprocity, ego will negatively evaluate candidates in order to shape their perception by the broader audience. The image of an aspiring seminar *participant* is illustrative. The participant may be outspoken in the critique of the speaker. While such opposition could harm the relationship with the speaker, it could also improve how others attending the seminar see the seminar participant as seeing through the flaws of someone else's work (Amabile 1983).

Research on social ties has shown that ties that are *positive* and *mutual* serve as an indicator of an ego's quality and trustworthiness (Gulati & Gargiulo, 1999; Podolny, 2001). By contrast, the evaluations we focus on are *negative* and *directed* (i.e., the candidate does not have to agree to be evaluated), and informs peers about ego's standards. The domain of negative ties is largely understudied (Labianca & Brass, 2006), and has typically focused on the recipient of negative ties (e.g., who becomes the victim of bullying or the subject of gossiping) (Rawlings & Friedkin, 2017). We, however, focus on the *sender* and show that the antecedents of negative ties can also be explained by strategic considerations in addition to structural explanations that have often been brought to the fore. The idea that ties can be both positive and negative is not new to the network literature and dates back to at least Simmel's seminal work (1964). Negative ties are rarer than positive ties with empirical research typically suggesting them to be less than 20 percent (Kane & Labianca, 2005; Labianca, Brass, & Gray, 1998). However, negative ties have greater effects on individuals' behavior than positive ties (Taylor, 1991). Given these disproportionate effects on behavior, it is surprising that so little research has been devoted to how they form and whether they form in different ways to positive ties (Wiseman & Duck, 1995). Recent research on social networks has made inroads into understanding how negative ties arise, investigating characteristics of the dyad. For instance, Rubineau, Lim, and Neblo (2019) find that "negative ties travel down

status hierarchies and target low status individuals, and a negative tie between two people becomes more likely as their status difference increases.” Similarly, Merluzzi (2017) studies negative ties in two US companies and finds that “women and men do not differ in their likelihood to cite a negative work tie, women are more likely (than men) to cite a woman as a negative tie.” These considerations can thus prevent us from forming productive ties as people are unlikely to seek advice from someone they dislike, even if that person is highly competent (Casciaro & Lobo, 2008). We offer a new perspective by looking at how negative ties arise from an evaluator’s strategic considerations. We find that prospective candidates adjust their current evaluations and are more likely to evaluate candidates negatively, but they refrain from opposing similar and highly engaged candidates and avoid elections in which the outcome is uncertain. In addition, this occurs in an environment stripped of many of the normally salient social cues available in a face-to-face environment. Our study also helps us understand the separation of clusters within network. We illustrate that ego has an incentive to evaluate people in their domain positively and candidates from other domains negatively. Over time – and if replicated by other peers – such behavior results in the alienation of groups and the fractionalization of networks.

A major strength of our study is that it enables us to analyze whether the behaviors we study result in the desired outcome of being elected in the future. In other words, do people form ties that are aligned with their objectives? Studies of network formation go to great lengths to understand how networks form, but it has been notoriously difficult to disentangle whether people form ties that are consistent with their objectives (Ingram & Morris, 2007). Our unique analysis contributes to this debate by suggesting that it is optimal to avoid evaluating candidates negatively when the outcome is uncertain, when ego and the candidate operate in the same domain as highly engaged candidates. Many negative evaluations decrease the likelihood of the approval of a request for

adminship, but at least *one* negative evaluation increases the odds. This suggests that peers cannot sprinkle positive evaluations and completely avoid negative evaluations, but have to occasionally do a negative evaluation so as to not appear sphinxlike.

Managerial implications

Our research also has important managerial implications. The goal of peer evaluations is to access and aggregate information about a candidate so as to reach a more comprehensive and objective evaluation. Our research shows how strategic behavior may undermine that goal. Our research allows us to speculate about how organizations may attenuate that problem. Organizations could make evaluations fully anonymous so that ego has no incentive to evaluate strategically. A less severe form could be to make ego's evaluations less accessible, for example, to make them available only with a delay or in a non-aggregated manner. Organizations could also limit permission to provide peer evaluations, for example, to people who have already been promoted. While we do not want to encourage this behavior, our post-hoc analysis also illustrates how actors may evaluate candidates strategically to foster their own promotion (i.e., provide negative evaluations, but not when the outcome of an election is uncertain or when the candidates operate in a similar domain).

Boundary conditions, limitations, and future research

Our study is subject to several boundary conditions. First, for a signal of negative ties to function, some of these interactions have to be made in the open. This is not always the case and data is scarce. Second, for ego to engage in more negative evaluations, a scope condition is that ego assumes the support they gather from the broad audience of peers by evaluating more candidates negatively outweighs the costs that ego is likely to encounter in the form of negative reciprocity. If this condition does *not* hold, negative ties are unlikely. We believe this condition to

hold if ego cannot control which peers will evaluate her. Ego thus needs to manage how they are perceived by the broad audience of peers as any one member of that audience may evaluate her. Third, the process of gaining authority obviously differs between communities and a firm. One important difference is that, unlike in a firm, there are not a fixed limited number of positions to which people can be elevated. Even if competition is more limited, as the situation allows many people to gain authority, and despite its social cost, people still express concerns about some of the candidates. This illustrates that the symbolic benefits of becoming a leader, as well as the benefits of gaining authority over collective work, is enough to lead to people forming negative ties. Inside a company, where stakes are higher and competition fiercer, the baseline rate of forming negative ties may be higher. Future research would do well to expand this research to other settings.

There are several limitations to our work, which pose opportunities for future research. We have suggested building on prior literature that negative evaluations help ego attract votes from peers because they signal high standards. Our data does not allow us to test these mechanisms directly. Ours is a special case, and there could be limited opportunities for generalization. Future work would do well to study positive and negative ties in tandem, but such data is notoriously difficult to come by (Labianca & Brass, 2006). An advantage of our theory is that there are few social cues on Wikipedia. While this provides a clean test of our theory, it is also unrealistic when observed in other situations where people have face-to-face interaction. More generally, research on evaluations is currently held back in secrecy. We hope that researchers get access to valuations and can uncover what shapes how actors evaluate one another.

FIGURES

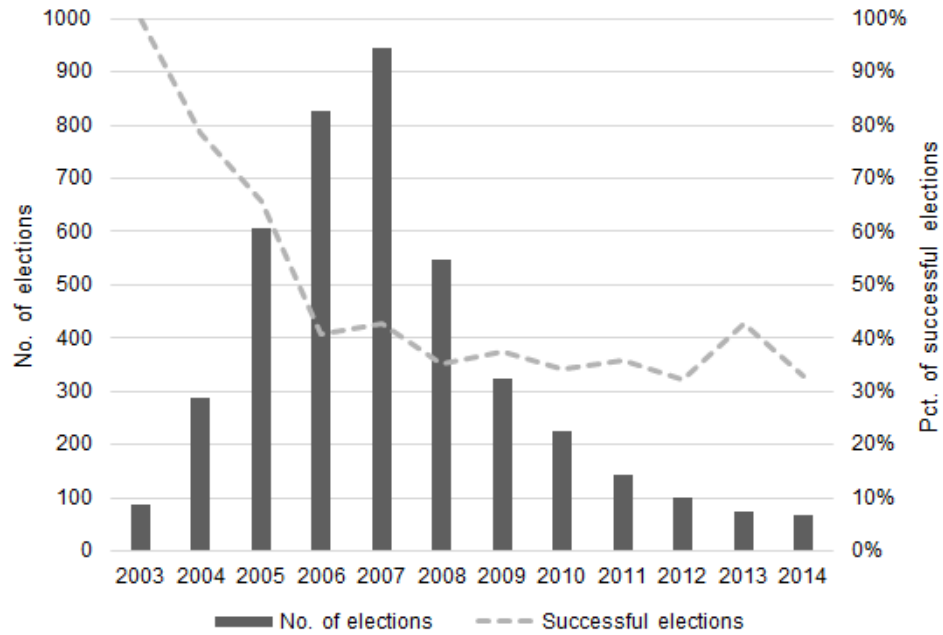


Figure 1. Number of elections and average number of participants over time

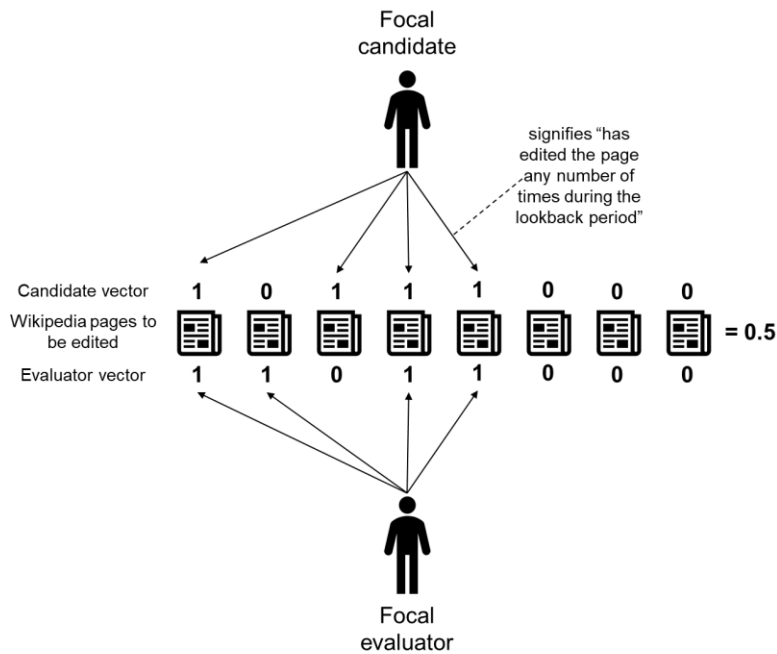


Figure 2. Construction of correlation score

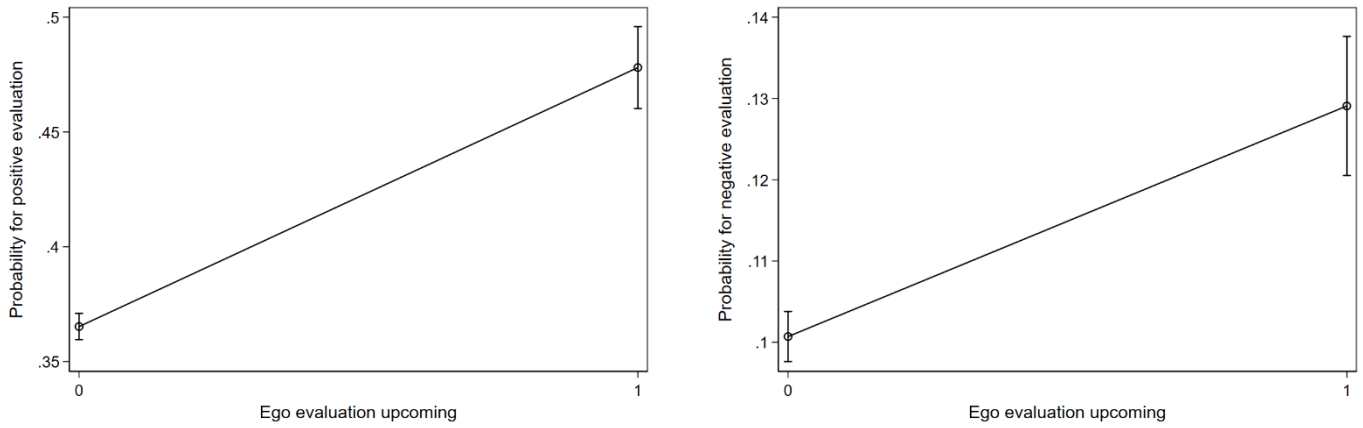


Figure 3: Predicted probability for positive (a) and negative (b) evaluation when ego is under evaluation.

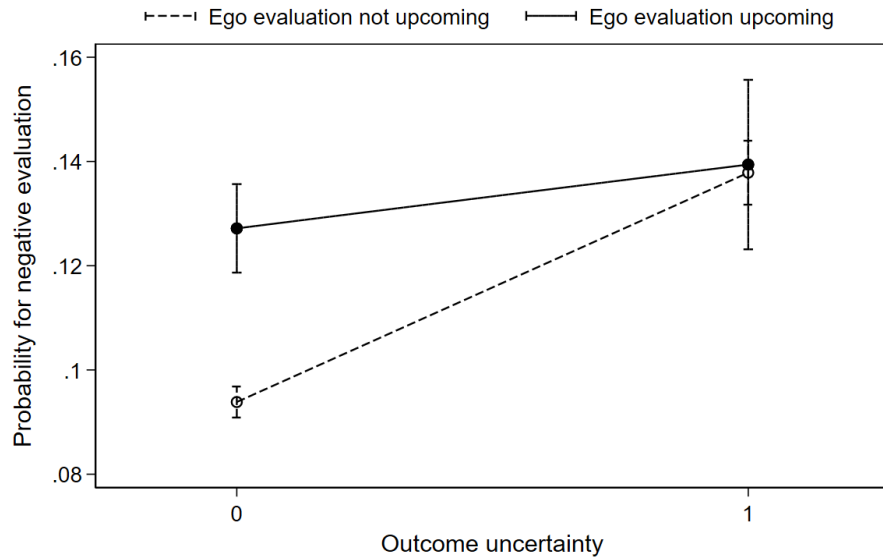


Figure 4: Predicted probability for negative evaluation when ego is under evaluation moderated by outcome uncertainty.

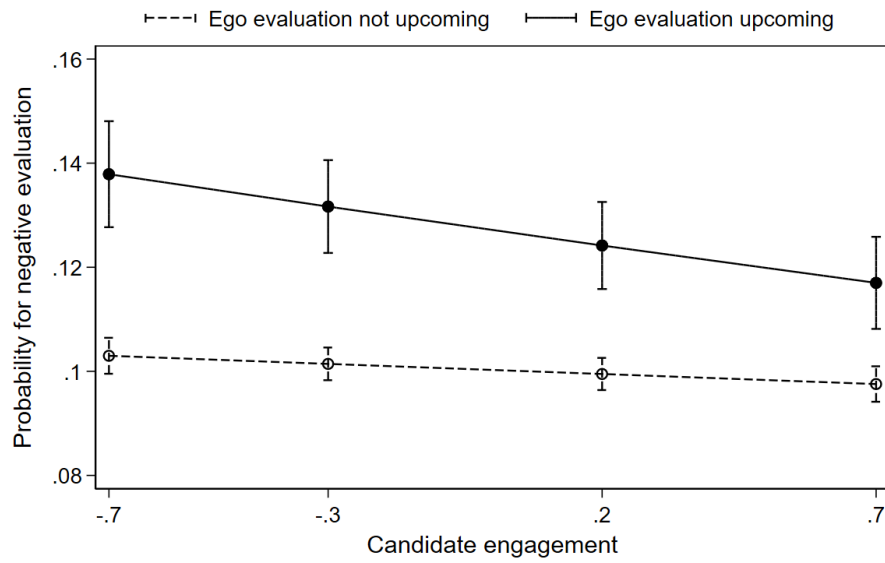


Figure 5: Predicted probability for negative evaluation when ego is under evaluation moderated by candidate engagement.

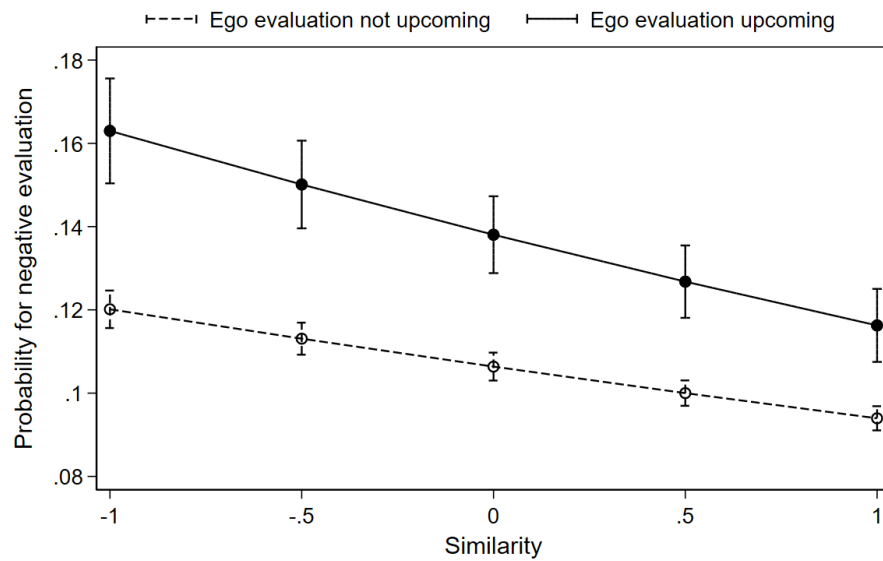


Figure 6: Predicted probability for negative evaluation when ego is under evaluation moderated by domain similarity.

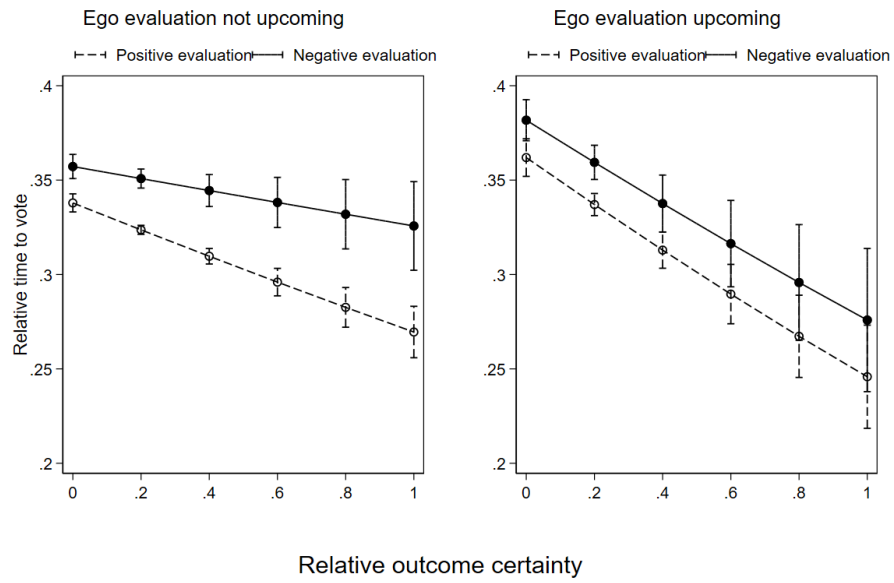


Figure 7: Predicted relative time to vote dependent on whether ego is under evaluation (right) or not (left) and type of evaluation (positive vs. negative)

TABLES

Table 1: Descriptive statistics with means and standard deviations

| | Dyads (N = 234,078) | Mean | S.D. | Min | Max | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|------|------------------------------|---------|---------|--------|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| (1) | Tendency to evaluate | 0.071 | 0.078 | 0.000 | 0.889 | 1 | | | | | | | | | |
| (2) | Engagement | 0.2 | 1.215 | -0.482 | 33.226 | 0.01 | 1 | | | | | | | | |
| (3) | Share of edits | 0.306 | 0.121 | 0.000 | 1.000 | 0.03 | 0.17 | 1 | | | | | | | |
| (4) | Tenure | 25.504 | 21.087 | 0.000 | 154.233 | -0.11 | -0.11 | 0.12 | 1 | | | | | | |
| (5) | Number of elec. participated | 69.494 | 99.98 | 0.000 | 962.000 | 0.52 | -0.01 | 0.01 | 0.31 | 1 | | | | | |
| (6) | Admin | 0.375 | 0.484 | 0.000 | 1.000 | 0.02 | 0.09 | -0.05 | 0.27 | 0.31 | 1 | | | | |
| (7) | Under evaluation | 0.154 | 0.360 | 0.000 | 1.000 | 0.05 | 0.00 | 0.05 | -0.19 | -0.12 | -0.33 | 1.00 | | | |
| (8) | Relative order to vote | 0.515 | 0.118 | 0.009 | 1.000 | -0.28 | -0.04 | -0.04 | 0 | -0.18 | -0.07 | -0.07 | 1 | | |
| (9) | Previous failed candidacy | 0.114 | 0.318 | 0.000 | 1.000 | 0.14 | -0.02 | 0.06 | -0.02 | 0.08 | -0.24 | 0.13 | -0.12 | 1 | |
| (10) | Tendency to reverse eval. | 0.78 | 1.911 | 0.000 | 29.000 | 0.31 | 0.01 | -0.02 | 0.04 | 0.33 | 0.13 | 0.04 | -0.19 | 0.05 | 1 |
| (11) | Deviation from the maj. vote | 0.212 | 0.112 | 0.000 | 1.000 | -0.08 | -0.09 | -0.1 | 0.04 | -0.09 | -0.13 | -0.05 | 0.05 | 0.05 | 0.01 |
| (12) | Prior evaluation | 0.711 | 0.453 | 0.000 | 1.000 | -0.04 | -0.03 | -0.04 | -0.02 | -0.01 | 0.02 | -0.02 | 0.02 | -0.01 | -0.01 |
| (13) | Edited pages | 7.288 | 1.293 | 0.000 | 10.924 | -0.02 | -0.06 | 0 | 0.05 | 0.02 | 0.02 | -0.03 | 0.02 | 0 | -0.01 |
| (14) | Engagement | 0.035 | 1.056 | -0.805 | 10.121 | -0.05 | 0.23 | 0.02 | -0.07 | -0.08 | -0.02 | 0 | 0.03 | -0.03 | -0.03 |
| (15) | Share of edits | 0.334 | 0.116 | 0.002 | 1.000 | 0.03 | -0.01 | 0.11 | 0.23 | 0.04 | 0 | -0.05 | -0.02 | 0 | 0 |
| (16) | Tenure | 25.154 | 21.081 | 0.000 | 119.833 | 0.04 | -0.13 | 0.15 | 0.51 | 0.1 | 0.02 | -0.11 | -0.02 | 0 | -0.01 |
| (17) | Edits commented | 0.921 | 0.14 | 0.000 | 1.000 | 0.02 | -0.16 | 0.02 | 0.16 | 0.07 | 0.03 | -0.05 | 0 | 0.01 | 0.01 |
| (18) | Edits reverted (Total) | 104.393 | 177.472 | 1.000 | 1899.000 | -0.02 | -0.05 | 0 | 0.02 | 0 | 0.01 | -0.02 | 0 | 0 | -0.01 |
| (19) | Uncertain outcome | 0.152 | 0.359 | 0.000 | 1.000 | -0.03 | -0.04 | -0.01 | 0.04 | 0 | 0.01 | -0.02 | 0.01 | 0 | -0.01 |
| (20) | Previous failed candidacy | 0.25 | 0.433 | 0.000 | 1.000 | -0.02 | -0.05 | -0.02 | 0.02 | 0.01 | 0.02 | -0.02 | 0 | 0.01 | 0 |
| (21) | Prior neg. evaluation | 0.003 | 0.057 | 0.000 | 1.000 | 0.01 | 0.01 | 0.01 | 0 | 0 | 0 | 0.01 | -0.02 | 0.09 | 0.01 |
| (22) | Prior pos. evaluation | 0.022 | 0.145 | 0.000 | 1.000 | 0 | 0.04 | 0 | 0 | 0 | 0.13 | -0.03 | -0.03 | 0.03 | 0.02 |
| (23) | Domain similarity | 0.435 | 1.448 | -1.596 | 66.315 | 0 | 0.08 | -0.18 | -0.07 | 0.02 | 0.08 | -0.02 | -0.04 | 0 | 0.04 |
| (24) | Edits of candidate reverted | 0.014 | 0.131 | 0.000 | 5.730 | -0.03 | 0.07 | 0.01 | -0.02 | -0.03 | 0 | -0.01 | 0 | 0 | -0.01 |
| (25) | Edits by candidate reverted | 0.017 | 0.143 | 0.000 | 5.707 | -0.03 | 0.09 | 0 | -0.01 | -0.02 | 0.03 | -0.01 | 0 | 0 | -0.01 |

| | (12) | (13) | (14) | (15) | (16) | (17) | (18) | (19) | (20) | (21) | (22) | (23) | (24) |
|----------------------------------|-------|------|-------|-------|-------|-------|------|------|------|------|------|------|------|
| (12) Prior evaluation | 1 | | | | | | | | | | | | |
| (13) Edited pages | 0.29 | 1 | | | | | | | | | | | |
| (14) Engagement | 0.1 | 0.53 | 1 | | | | | | | | | | |
| (15) Share of edits | -0.12 | 0.24 | 0.15 | 1 | | | | | | | | | |
| (16) Tenure | -0.1 | 0.02 | -0.12 | 0.23 | 1 | | | | | | | | |
| (17) Edits commented | 0.22 | 0.29 | -0.09 | -0.02 | 0.15 | 1 | | | | | | | |
| (18) Edits reverted (Total) | 0.13 | 0.53 | 0.48 | 0 | -0.06 | 0.11 | 1 | | | | | | |
| (19) Uncertain outcome | 0.06 | 0.06 | 0.01 | -0.02 | 0.08 | 0.01 | 0.07 | 1 | | | | | |
| (20) Previous failed candidacy | 0.17 | 0.13 | 0.07 | -0.04 | 0.08 | 0.03 | 0.16 | 0.12 | 1 | | | | |
| (21) Prior neg. evaluation | 0.04 | 0.02 | 0 | -0.01 | 0 | 0.01 | 0 | 0 | 0.03 | 1 | | | |
| (22) Prior pos. evaluation | 0.09 | 0.05 | 0.02 | -0.01 | -0.01 | 0.03 | 0.02 | 0.01 | 0.06 | 0.02 | 1 | | |
| (23) Domain similarity | 0.23 | 0.08 | 0.1 | -0.22 | -0.12 | 0.06 | 0.09 | 0.03 | 0.1 | 0.05 | 0.15 | 1 | |
| (24) Edits of candidate reverted | 0.02 | 0.05 | 0.11 | 0.01 | -0.02 | 0 | 0.04 | 0 | 0.01 | 0.02 | 0.03 | 0.21 | 1 |
| (25) Edits of candidate reverted | 0.02 | 0.05 | 0.09 | 0 | -0.03 | -0.02 | 0.08 | 0 | 0.02 | 0.01 | 0.03 | 0.22 | 0.42 |

| Election (N = 3,448) | Mean | S.D. | Min | Max | | | | | | |
|----------------------|--------|--------|-------|-------|------|-------|-------|------|------|--|
| Support Votes | 37.312 | 38.922 | 0 | 256 | 1.00 | | | | | |
| Oppose Votes | 10.272 | 12.685 | 0 | 124 | 0.08 | 1.00 | | | | |
| Neutral Votes | 2.981 | 3.955 | 0 | 44 | 0.14 | 0.68 | 1.00 | | | |
| No of Voters | 50.565 | 43.376 | 1 | 383 | 0.93 | 0.43 | 0.41 | 1.00 | | |
| Successful Election | 0.456 | 0.498 | 0.000 | 1.000 | 0.63 | -0.44 | -0.25 | 0.41 | 1.00 | |

Table 2: Multinomial logistic regression, DV: Decision whether and how to evaluate (no evaluation as base case). Moderators included

| | 1.1 (H1) | | 1.2 (H2) | | 1.3 (H3) | | 1.4 (H4) | | 1.5 (Comb.) | |
|----------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| | Negative | Positive | Negative | Positive | Negative | Positive | Negative | Positive | Negative | Positive |
| Under evaluation | 0.586*** (0.059) | 0.585*** (0.048) | 0.656*** (0.060) | 0.596*** (0.048) | 0.564*** (0.059) | 0.584*** (0.048) | 0.599*** (0.060) | 0.576*** (0.050) | 0.643*** (0.060) | 0.587*** (0.050) |
| Uncertain outcome | 0.415*** (0.028) | -0.077*** (0.019) | 0.466*** (0.028) | -0.069*** (0.020) | 0.414*** (0.028) | -0.077*** (0.019) | 0.414*** (0.028) | -0.076*** (0.019) | 0.463*** (0.028) | -0.069*** (0.020) |
| Engagement | -0.086*** (0.014) | -0.071*** (0.010) | -0.087*** (0.014) | -0.071*** (0.010) | -0.070*** (0.015) | -0.071*** (0.010) | -0.086*** (0.014) | -0.071*** (0.010) | -0.073*** (0.015) | -0.071*** (0.010) |
| Domain similarity | -0.163*** (0.009) | -0.009 (0.007) | -0.163*** (0.009) | -0.009 (0.007) | -0.163*** (0.009) | -0.009 (0.007) | -0.155*** (0.010) | -0.011 (0.007) | -0.158*** (0.010) | -0.011 (0.007) |
| Under evaluation * | | | -0.438*** (0.084) | -0.083 (0.060) | | | | | -0.414*** (0.083) | -0.083 (0.060) |
| Uncertain outcome | | | | | | | | | | |
| Under evaluation * | | | | | -0.143*** (0.038) | -0.002 (0.027) | | | -0.118** (0.036) | -0.005 (0.027) |
| Engagement | | | | | | | | | | |
| Under evaluation * | | | | | | | -0.087** (0.032) | 0.017 (0.026) | -0.063* (0.031) | 0.019 (0.026) |
| Domain similarity | | | | | | | | | | |
| Ego-level controls: | | | | | | | | | | |
| Tendency to evaluate | 0.359 (0.362) | -0.645* (0.301) | 0.353 (0.362) | -0.645* (0.301) | 0.344 (0.362) | -0.644* (0.301) | 0.348 (0.361) | -0.642* (0.302) | 0.334 (0.361) | -0.643* (0.302) |
| Edited pages | 0.076*** (0.015) | -0.007 (0.013) | 0.077*** (0.015) | -0.007 (0.013) | 0.076*** (0.015) | -0.007 (0.013) | 0.077*** (0.015) | -0.007 (0.013) | 0.077*** (0.015) | -0.007 (0.013) |
| Engagement | -0.029 (0.016) | -0.010 (0.012) | -0.028 (0.016) | -0.010 (0.012) | -0.030 (0.015) | -0.010 (0.012) | -0.029 (0.015) | -0.010 (0.012) | -0.030 (0.015) | -0.010 (0.012) |
| Share of edits | -0.804*** (0.205) | 0.255 (0.210) | -0.807*** (0.204) | 0.255 (0.210) | -0.802*** (0.204) | 0.255 (0.210) | -0.810*** (0.204) | 0.256 (0.210) | -0.810*** (0.204) | 0.255 (0.210) |
| Tenure | 0.006*** (0.001) | 0.001 (0.001) | 0.006*** (0.001) | 0.001 (0.001) | 0.006*** (0.001) | 0.001 (0.001) | 0.006*** (0.001) | 0.001 (0.001) | 0.006*** (0.001) | 0.001 (0.001) |
| Number of elections participated | 0.000 (0.000) | 0.001* (0.000) | 0.000 (0.000) | 0.001* (0.000) | 0.000 (0.000) | 0.001* (0.000) | 0.000 (0.000) | 0.001* (0.000) | 0.000 (0.000) | 0.001* (0.000) |
| Admin | 0.056 (0.049) | -0.116** (0.036) | 0.055 (0.049) | -0.116** (0.036) | 0.057 (0.049) | -0.116** (0.036) | 0.054 (0.049) | -0.115** (0.037) | 0.055 (0.049) | -0.115** (0.037) |

| | | | | | | | | | | |
|-------------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Relative order to vote | 0.049 (0.132) | 0.073 (0.099) | 0.050 (0.132) | 0.072 (0.099) | 0.055 (0.132) | 0.072 (0.099) | 0.050 (0.132) | 0.072 (0.099) | 0.056 (0.132) | 0.072 (0.099) |
| Previous failed candidacy | -0.248*** (0.065) | -0.327*** (0.052) | -0.249*** (0.065) | -0.327*** (0.052) | -0.250*** (0.065) | -0.327*** (0.052) | -0.246*** (0.065) | -0.328*** (0.052) | -0.248*** (0.064) | -0.328*** (0.052) |
| Tendency to reverse evaluations | 0.018 (0.016) | 0.015* (0.006) | 0.019 (0.016) | 0.015* (0.006) | 0.018 (0.016) | 0.015* (0.006) | 0.018 (0.016) | 0.015* (0.006) | 0.019 (0.016) | 0.015* (0.006) |
| Deviation from the majority vote | 4.784*** (0.139) | -1.792*** (0.104) | 4.783*** (0.139) | -1.793*** (0.104) | 4.781*** (0.139) | -1.793*** (0.104) | 4.783*** (0.139) | -1.792*** (0.104) | 4.779*** (0.139) | -1.793*** (0.104) |
| Candidate-level controls: | | | | | | | | | | |
| Prior evaluation | -0.324*** (0.021) | -0.027 (0.014) | -0.324*** (0.021) | -0.027 (0.014) | -0.324*** (0.021) | -0.027 (0.014) | -0.322*** (0.021) | -0.027 (0.014) | -0.322*** (0.021) | -0.027 (0.014) |
| Edited pages | -0.222*** (0.010) | 0.197*** (0.008) | -0.222*** (0.010) | 0.196*** (0.008) | -0.221*** (0.010) | 0.196*** (0.008) | -0.222*** (0.010) | 0.196*** (0.008) | -0.221*** (0.010) | 0.196*** (0.008) |
| Share of edits | 0.026 (0.087) | -0.009 (0.055) | 0.028 (0.087) | -0.009 (0.055) | 0.027 (0.087) | -0.009 (0.055) | 0.022 (0.087) | -0.008 (0.055) | 0.025 (0.087) | -0.008 (0.055) |
| Tenure | -0.012*** (0.001) | 0.006*** (0.000) | -0.012*** (0.001) | 0.006*** (0.000) | -0.012*** (0.001) | 0.006*** (0.000) | -0.012*** (0.001) | 0.006*** (0.000) | -0.012*** (0.001) | 0.006*** (0.000) |
| Edits commented | -1.479*** (0.066) | 0.793*** (0.056) | -1.476*** (0.066) | 0.793*** (0.056) | -1.478*** (0.066) | 0.794*** (0.056) | -1.475*** (0.066) | 0.792*** (0.056) | -1.473*** (0.066) | 0.792*** (0.056) |
| Edits reverted (Total) | 0.001*** (0.000) | -0.001*** (0.000) | 0.001*** (0.000) | -0.001*** (0.000) | 0.001*** (0.000) | -0.001*** (0.000) | 0.001*** (0.000) | -0.001*** (0.000) | 0.001*** (0.000) | -0.001*** (0.000) |
| Previous failed candidacy | 0.422*** (0.019) | -0.162*** (0.015) | 0.422*** (0.019) | -0.162*** (0.015) | 0.422*** (0.019) | -0.162*** (0.015) | 0.422*** (0.019) | -0.163*** (0.015) | 0.422*** (0.019) | -0.163*** (0.015) |
| Dyad-level controls: | | | | | | | | | | |
| Prior neg. evaluation | 0.519*** (0.150) | 0.053 (0.122) | 0.517*** (0.150) | 0.052 (0.122) | 0.522*** (0.150) | 0.053 (0.122) | 0.523*** (0.151) | 0.051 (0.122) | 0.523*** (0.151) | 0.051 (0.122) |
| Prior pos. evaluation | 0.029 (0.073) | 0.617*** (0.052) | 0.028 (0.073) | 0.617*** (0.052) | 0.027 (0.073) | 0.617*** (0.052) | 0.027 (0.073) | 0.619*** (0.052) | 0.025 (0.073) | 0.619*** (0.052) |
| Edits of candidate reverted | 0.725*** (0.073) | 0.559*** (0.052) | 0.723*** (0.073) | 0.559*** (0.052) | 0.722*** (0.073) | 0.560*** (0.052) | 0.724*** (0.073) | 0.559*** (0.052) | 0.721*** (0.073) | 0.559*** (0.052) |

| | | | | | | | | | | |
|---|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Edits by candidate reverted | (0.098) 0.839*** | (0.088) 0.518*** | (0.098) 0.837*** | (0.088) 0.517*** | (0.098) 0.834*** | (0.088) 0.517*** | (0.098) 0.835*** | (0.088) 0.518*** | (0.098) 0.831*** | (0.088) 0.517*** |
| Edits of candidate reverted * Edits by candidate reverted | (0.103) -0.175*** | (0.073) -0.401*** | (0.103) -0.174*** | (0.073) -0.401*** | (0.103) -0.175*** | (0.073) -0.401*** | (0.103) -0.178*** | (0.073) -0.400*** | (0.103) -0.176*** | (0.073) -0.400*** |
| Constant | (0.047) -0.351* | (0.068) -1.781*** | (0.048) -0.366* | (0.068) -1.782*** | (0.047) -0.348* | (0.068) -1.779*** | (0.048) -0.360* | (0.068) -1.775*** | (0.048) -0.370* | (0.068) -1.775*** |
| Year fixed effects | (0.174) Yes | (0.152) Yes | (0.174) Yes | (0.152) Yes | (0.174) Yes | (0.152) Yes | (0.173) Yes | (0.152) Yes | (0.174) Yes | (0.152) Yes |
| R-squared | 0.239 | | 0.239 | | 0.239 | | 0.239 | | 0.239 | |
| N | 234,144 | | 234,144 | | 234,144 | | 234,144 | | 234,144 | |

Robust standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Logistic regression of approval rating (for request of being promoted to admin) dependent on main independent variables.

| | 2.1 | 2.2 | 2.3 | 2.4 | 2.5 |
|----------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Voting tactics: | | | | | |
| Pos. eval. overall (H1a) | 0.001*** (0.000) | 0.001** (0.000) | 0.001 (0.000) | 0.001 (0.000) | 0.000 (0.000) |
| Neg. eval. overall (H1b) | -0.002* (0.001) | -0.003** (0.001) | -0.001 (0.001) | -0.002 (0.001) | -0.002 (0.001) |
| Neg. eval. dummy (H1b) | | 0.037*** (0.013) | 0.021 (0.015) | 0.006 (0.016) | 0.020 (0.016) |
| Neg. eval. close elections (H2b) | | | -0.147*** (0.051) | | |
| Neg. eval. eng. candidates (H3b) | | | | -0.109** (0.054) | |
| Neg. eval. sim. candidates (H4b) | | | | | -0.099** (0.044) |
| Candidate-level controls: | | | | | |
| Edited pages | 0.099*** (0.006) | 0.097*** (0.006) | 0.094*** (0.008) | 0.094*** (0.009) | 0.090*** (0.008) |
| Pct. Edits commented | 0.591*** (0.051) | 0.578*** (0.050) | 0.636*** (0.057) | 0.546*** (0.061) | 0.592*** (0.064) |
| Reverts ratio | -0.135** (0.060) | -0.137** (0.060) | -0.117* (0.065) | -0.077 (0.071) | -0.127* (0.066) |
| Reverts total | -0.000*** (0.000) | -0.000*** (0.000) | -0.000*** (0.000) | -0.000*** (0.000) | -0.000*** (0.000) |
| Tenure | 0.004*** (0.001) | 0.005*** (0.001) | 0.004*** (0.001) | 0.004*** (0.001) | 0.004*** (0.001) |
| Constant | -0.304*** (0.056) | -0.297*** (0.055) | -0.362*** (0.071) | -0.247*** (0.076) | -0.265*** (0.074) |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes |
| Repeated Candidacy Dummies | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.364 | 0.366 | 0.285 | 0.268 | 0.299 |
| N | 1,892 | 1,892 | 1,516 | 1,261 | 1,363 |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

BIBLIOGRAPHY

- Aadland, E., Cattani, G., & Ferriani, S. 2018a. The Social Structure of Consecration in Cultural Fields: The Influence of Status and Social Distance in Audience–Candidate Evaluative Processes. *Frontiers of Creative Industries: Exploring Structural and Categorical Dynamics*, 55: 129-157.
- Aadland, E., Cattani, G., & Ferriani, S. 2018b. Friends, Cliques and Gifts: Social Proximity and Recognition in Peer-Based Tournament Rituals. *Academy of Management Journal*, in press.
- Amabile, T. M. 1983. The Social-Psychology of Creativity - a Componential Conceptualization. *Journal of Personality and Social Psychology*, 45(2): 357-376.
- Balla, S. J., Lawrence, E. D., Maltzman, F., & Sigelman, L. 2002. Partisanship, Blame Avoidance, and the Distribution of Legislative Pork. *American Journal of Political Science*, 46(3): 515-525.
- Beljean, S., Chong, P., & Lamont, M. 2016. A Post-Bourdieuian Sociology of Valuation and Evaluation for the Field of Cultural Production. In L. Hanquinet & M. Savage (Eds.), *Routledge international handbook of the sociology of art and culture*: 38-48.
- Blank, G. 2006. *Critics, Ratings, and Society: The Sociology of Reviews*. Plymouth, UK: Rowman & Littlefield Publishers.
- Bolino, M. C., Kacmar, K. M., Turnley, W. H., & Gilstrap, J. B. 2008. A Multi-Level Review of Impression Management Motives and Behaviors. *Journal of Management*, 34(6): 1080-1109.
- Boudreau, K. J., Guinan, E. C., Lakhani, K. R., & Riedl, C. 2016. Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science. *Manage Sci*, 62(10): 2765-2783.
- Bourdieu, P. 1993. *The Field of Cultural Production: Essays on Art and Literature*. US: Columbia University Press.
- Bowers, A. 2015. Relative Comparison and Category Membership: The Case of Equity Analysts. *Organization Science*, 26(2): 571-583.
- Bowers, A. 2019. Balanced but not Fair: Strategic Balancing, Rating Allocations, and Third-party Intermediaries. *Strategic Organization*: DOI: 1476127019828359.
- Bozeman, D. P. & Kacmar, K. M. 1997. A Cybernetic Model of Impression Management Processes in Organizations. *Organizational Behavior and Human Decision Processes*, 69(1): 9-30.
- Burt, Ronald S. 2004. Structural Holes and Good Ideas. *American Journal of Sociology*, 110(2): 349-399.
- Casciaro, T. 1998. Seeing Things Clearly: Social Structure, Personality, and Accuracy in Social Network Perception. *Social Networks*, 20(4): 331-351.
- Casciaro, T. & Lobo, M. S. 2008. When Competence is Irrelevant: The Role of Interpersonal Affect in Task-Related Ties. *Administrative Science Quarterly*, 53(4): 655-684.
- Cattani, G., Ferriani, S., & Allison, P. D. 2014. Insiders, Outsiders, and the Struggle for Consecration in Cultural Fields. *American Sociological Review*, 79(2): 258-281.
- Chen, D. L., Moskowitz, T. J., & Shue, K. 2016. Decision Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires. *Quarterly Journal of Economics*, 131(3): 1181-1242.
- Chong, P. 2013. Legitimate Judgment in Art, the Scientific World Reversed? Maintaining Critical Distance in Evaluation. *Social Studies of Science*, 43(2): 265-281.
- Chong, P. 2015. Playing nice, being mean, and the space in between: Book critics and the difficulties of writing bad reviews. In A. B. Antal, M. Hutter, & D. Stark (Eds.), *Moments of Valuation: Exploring Sites of Dissonance*. : 133-146. Oxford, UK: Oxford University Press.
- Correll, S. J., Ridgeway, C. L., Zuckerman, E. W., Jank, S., Jordan-Bloch, S., & Nakagawa, S. 2017. It's the Conventional Thought That Counts: How Third-Order Inference Produces Status Advantage. *American Sociological Review*, 82(2): 297-327.
- Criscuolo, P., Dahlander, L., Grohsjean, T., & Salter, A. 2017. Evaluating Novelty: The Role of Panels in the Selection of R&D Projects. *Academy of Management Journal*, 60(2): 433-460.

- Ellwardt, L., Labianca, G., & Wittek, R. 2012. Who are the objects of positive and negative gossip at work?: A social network perspective on workplace gossip. *Social Networks*, 34(2): 193-205.
- Ertug, G. & Castellucci, F. 2014. Who shall get more? How intangible assets and aspiration levels affect the valuation of resource providers. *Strategic Organization*, 13(1): 6-31.
- Fernandez-Mateo, I. & Coh, M. 2015. Coming with Baggage: Past Rejections and the Evolution of Market Relationships. *Organization Science*, 26(5): 1381-1399.
- Filippas, A., Horton, J. J., & Golden, J. 2018. *Reputation Inflation*. Paper presented at the Proceedings of the 2018 ACM Conference on Economics and Computation, Ithaca, NY, USA.
- Flynn, F. J. 2006. How Much is it Worth to You? Subjective Evaluations of Help in Organizations. *Research in Organizational Behavior*, 27: 133-174.
- Gagne, F., Khan, A., Lydon, J., & To, M. 2008. When Flattery Gets You Nowhere: Discounting Positive Feedback as a Relationship Maintenance Strategy. *Canadian Journal of Behavioural Science-Revue Canadienne Des Sciences Du Comportement*, 40(2): 59-68.
- Gallus, J. 2017. Fostering Public Good Contributions with Symbolic Awards: A Large-Scale Natural Field Experiment at Wikipedia. *Management Science*, 63(12): 3999-4015.
- Gargiulo, M. 1993. Two-Step Leverage: Managing Constraint in Organizational Politics. *Administrative Science Quarterly*, 38(1): 1-19.
- Gargiulo, M., Ertug, G., & Galunic, C. 2009. The Two Faces of Control: Network Closure and Individual Performance among Knowledge Workers. *Administrative Science Quarterly*, 54(2): 299-333.
- Gouldner, A. W. 1960. The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review*, 25(2): 161-178.
- Green, E. A. & Daniels, D. P. 2014. What does it take to call a strike? Three biases in umpire decision making, *MIT Sloan Sports Analytics Conference*. Cambridge, MA, USA: MIT Sloan.
- Gulati, R. & Gargiulo, M. 1999. Where Do Interorganizational Networks Come From? *American Journal of Sociology*, 104(5): 1439-1493.
- Gulati, R. & Higgins, M. C. 2003. Which ties matter when? The contingent effects of interorganizational partnerships on IPO success. *Strategic Management Journal*, 24(2): 127-144.
- Hallen, B. L. 2008. The Causes and Consequences of the Initial Network Positions of New Organizations: From Whom Do Entrepreneurs Receive Investments? *Administrative Science Quarterly*, 53(4): 685-718.
- Higgins, C. A., Judge, T. A., & Ferris, G. R. 2003. Influence tactics and work outcomes: a meta-analysis. *Journal of Organizational Behavior*, 24(1): 89-106.
- Howard, C. D., Barrett, A. F., & Frick, T. W. 2010. Anonymity to Promote Peer Feedback: Pre-Service Teachers' Comments in Asynchronous Computer-Mediated Communication. *Journal of Educational Computing Research*, 43(1): 89-112.
- Ingram, P. & Morris, M. W. 2007. Do People Mix at Mixers? Structure, Homophily, and the "Life of the Party". *Administrative Science Quarterly*, 52(4): 558-585.
- Jemielniak, D. 2014. *Common Knowledge? An Ethnography of Wikipedia*. Stanford: Stanford University Press.
- Jonczyk, C. D., Lee, Y. G., Galunic, C. D., & Bensaou, B. M. 2016. Relational Changes during Role Transitions: The Interplay of Efficiency and Cohesion. *Academy of Management Journal*, 59(3): 956-982.
- Kane, G. & Labianca, G. 2005. Accounting for clergy's social ledgers: Mixed blessings associated with direct and indirect negative ties in a religious organization, *Intra-Organizational Network (ION) Conference, Atlanta*.
- Kim, J. W. & King, B. G. 2014. Seeing Stars: Matthew Effects and Status Bias in Major League Baseball Umpiring. *Management Science*, 60(11): 2619-2644.

- Kittur, A., Suh, B., Pendleton, B. A., & Chi, E. H. 2007. He says, she says: conflict and coordination in Wikipedia. In M. B. Rosson & D. Gilmore (Eds.), *SIGCHI conference on Human factors in computing systems*: 453-462. San Jose, CA, USA: ACM.
- Klapper, H. & Reitzig, M. 2018. On the effects of authority on peer motivation: Learning from Wikipedia. *Strategic Management Journal*, 39(8): 2178-2203.
- Kleinbaum, A. M. 2012. Organizational Misfits and the Origins of Brokerage in Intrafirm Networks. *Administrative Science Quarterly*, 57(3): 407-452.
- Krackhardt, D. 1990. Assessing the Political Landscape: Structure, Cognition, and Power in Organizations. *Administrative Science Quarterly*, 35(2): 342-369.
- Labianca, G., Brass, D. J., & Gray, B. 1998. Social networks and perceptions of intergroup conflict: The role of negative relationships and third parties. *Academy of Management Journal*, 41(1): 55-67.
- Labianca, G. & Brass, D. J. 2006. Exploring the social ledger: Negative relationships and negative asymmetry in social networks in organizations. *Academy of Management Review*, 31(3): 596-614.
- Lamont, M. 2012. Toward a Comparative Sociology of Valuation and Evaluation. *Annual Review of Sociology*, 38(1): 201-221.
- Lee, G. K. & Cole, R. E. 2003. From a firm-based to a community-based model of knowledge creation: The case of the Linux kernel development. *Organization Science*, 14(6): 633-649.
- Leskovec, J., Huttenlocher, D. P., & Kleinberg, J. M. 2010. Governance in social media: A case study of the wikipedia promotion process, *ICWSM*: 98-105.
- Merluzzi, J. 2017. Gender and Negative Network Ties: Exploring Difficult Work Relationships Within and Across Gender. *Organization Science*, 28(4): 636-652.
- Metiu, A. & Kogut, B. 2001. Open-Source Software Development and Distributed Innovation. *Oxford Review of Economic Policy*, 17(2): 248-264.
- Nagaraj, A. 2018. Does Copyright Affect Reuse? Evidence from Google Books and Wikipedia. *Management Science*, 64(7): 3091-3107.
- O'Mahony, S. & Ferraro, F. 2007. The emergence of governance in an open source community. *Academy of Management Journal*, 50(5): 1079-1106.
- Piskorski, M. J. & Gorbatai, A. 2017. Testing Coleman's Social-Norm Enforcement Mechanism: Evidence from Wikipedia. *American Journal of Sociology*, 122(4): 1183-1222.
- Podolny, J. M. 1993. A Status-Based Model of Market Competition. *American Journal of Sociology*, 98(4): 829-872.
- Podolny, Joel M. 2001. Networks as the Pipes and Prisms of the Market. *American Journal of Sociology*, 107(1): 33-60.
- Puranam, P., Alexy, O., & Reitzig, M. 2014. What's "new" about new forms of organizing? *Academy of Management Review*, 39(2): 162-180.
- Rawlings, C. M. & Friedkin, N. E. 2017. The Structural Balance Theory of Sentiment Networks: Elaboration and Test. *American Journal of Sociology*, 123(2): 510-548.
- Ridgeway, C. L. & Correll, S. J. 2006. Consensus and the creation of status beliefs. *Social Forces*, 85(1): 431-453.
- Rosenbaum, P. R. & Rubin, D. B. 1984. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association*, 79(387): 516-524.
- Rossmann, G., Esparza, N., & Bonacich, P. 2010. I'd Like to Thank the Academy, Team Spillovers, and Network Centrality. *American Sociological Review*, 75(1): 31-51.
- Rubineau, B., Lim, Y., & Neblo, M. 2019. Low status rejection: How status hierarchies influence negative tie formation. *Social Networks*, 56: 33-44.
- Salganik, M. J., Dodds, P. S., & Watts, D. J. 2006. Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. *Science*, 311(5762): 854-856.

- Shymko, Y. & Roulet, T. J. 2017. When Does Medici Hurt Da Vinci? Mitigating the Signaling Effect of Extraneous Stakeholder Relationships in the Field of Cultural Production. *Academy of Management Journal*, 60(4): 1307-1338.
- Sorenson, O. & Stuart, Toby E. 2001. Syndication Networks and the Spatial Distribution of Venture Capital Investments. *American Journal of Sociology*, 106(6): 1546-1588.
- Srivastava, S. B., Goldberg, A., Manian, V. G., & Potts, C. 2018. Enculturation Trajectories: Language, Cultural Adaptation, and Individual Outcomes in Organizations. *Management Science*, 64(3): 1348-1364.
- Stuart, T. E., Hoang, H., & Hybels, R. C. 1999. Interorganizational endorsements and the performance of entrepreneurial ventures. *Administrative Science Quarterly*, 44(2): 315-349.
- Taggar, S. & Brown, T. C. 2006. Interpersonal Affect and Peer Rating Bias in Teams. *Small Group Research*, 37(1): 86-111.
- Taylor, S. E. 1991. Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin*, 110(1): 67-85.
- Trapido, D. 2013. Dual Signals: How Competition Makes or Breaks Interfirm Social Ties. *Organization Science*, 24(2): 498-512.
- Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. 2015. What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers & Education*, 81: 123-132.
- Vis, B. & van Kersbergen, K. 2016. Why and how do Political Actors Pursue Risky Reforms? *Journal of Theoretical Politics*, 19(2): 153-172.
- Weaver, R. K. 1986. The Politics of Blame Avoidance. *Journal of Public Policy*, 6(4): 371-398.
- Wiseman, J. P. & Duck, S. 1995. Having and managing enemies: A very challenging relationship. In S. Duck & J. T. Wood (Eds.), *Confronting relationship challenges.*: 43-72. Thousand Oaks, CA, US: Sage Publications, Inc.
- Zhang, X. & Zhu, F. 2011. Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review*, 101(4): 1601-1615.
- Zuckerman, E. W. 2004. Towards the Social Reconstruction of an Interdisciplinary Turf War. *American Sociological Review*, 69(3): 458-465.