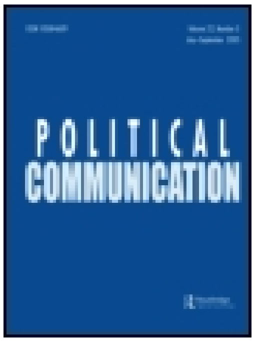

Political Communication

ISSN: 1058-4609 (Print) 1091-7675 (Online) Journal homepage: http://www.tandfonline.com/loi/upcp20

Political Fact-Checking on Twitter: When Do Corrections Have an Effect?

Drew B. Margolin, Aniko Hannak & Ingmar Weber

Political Communication, 00:1–24, 2017

ISSN: 1058-4609 print / 1091-7675 online
DOI: https://doi.org/10.1080/10584609.2017.1334018

Political Fact-Checking on Twitter: When Do Corrections Have an Effect?

DREW B. MARGOLIN, ANIKO HANNAK, and INGMAR WEBER

Research suggests that fact checking corrections have only a limited impact on the spread of false rumors. However, research has not considered that fact-checking may be socially contingent, meaning there are social contexts in which truth may be more or less preferred. In particular, we argue that strong social connections between fact-checkers and rumor spreaders encourage the latter to prefer sharing accurate information, making them more likely to accept corrections. We test this argument on real corrections made on Twitter between Janurary 2012 and April, 2014. As hypothesized, we find that individuals who follow and are followed by the people who correct them are significantly more likely to accept the correction than individuals confronted by strangers. We then replicate our findings on new data drawn from November 2015 to February, 2016. These findings suggest that the underlying social structure is an important factor in the correction of misinformation.

Keywords accountability, fact-checking, misinformation, rumor, social networks

Introduction

The dissemination and acceptance of false political rumors threatens the efficacy of democracy (Allport & Postman, 1947; Conover, Gonçalves, Flammini, & Menczer, 2012; Gottfried, Hardy, Winneg, & Jamieson, 2013). The dissemination of fake news, in particular, has received renewed attention in the news media after the discovery of its proliferation during the 2016 U.S. presidential election (Giglietto, Iannelli, Rossi, & Valeriani, 2016). Yet, unfortunately, empirical studies of rumor and misinformation suggest that stemming the spread of fake news is difficult. Experimental research suggests that when misinformation is consistent with an individual's existing worldview corrections will have a minimal impact (Einwiller & Kamins, 2008; Garrett, Nisbet, & Lynch, 2013; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012) and in some cases, cause the individual to hold more strongly to false beliefs (Nyhan & Reifler, 2010). Recent large-scale studies of rumor diffusion are equally discouraging. In a national survey, Garrett (2011) finds that exposure to corrections had only a small impact on belief. Friggeri, Adamic, Eckles, and Cheng (2014) show that rumor propagation on Facebook is slowed

Drew B. Margolin is Assistant Professor, Department of Communication, Cornell University. Aniko Hannak is Doctoral Candidate, College of Computing and Information Science, Northeastern University. Ingmar Weber is a Senior Scientist, Qatar Computing Institute, Hamad Bin Khalifa University.

Address correspondence to Drew Margolin Department of Communication, 472 Mann Library Building, Cornell University, Ithaca NY 14853, USA. E-mail: dm658@cornell.edu

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/UPCP.

by the presentation of correcting information; however, they also find that many rumors can be corrected repeatedly and still continue to diffuse through a social network. Shin, Driscoll, and Bar (2016) find that the publication of research debunking a rumor by a fact-checking website does not substantially alter the rate of rumor adoption in Twitter. Analyses of large-scale information consumption patterns on social media also suggest that some people may seek out rumors and avoid sources that might correct them (Bessi et al., 2015; Mocanu, Rossi, Zhang, Karsai, & Quattrociocchi, 2015). While some studies do show that corrections can have an effect (Weeks, 2015) the effects are often weak and require special contexts.

These findings can be discouraging for those who hope that individuals will internalize and practice deliberative norms in political communication (Mutz, 2008). According to the norms of deliberative democracy, individuals should prefer factual information, forming their views about policy and other public views like *intuitive scientists* (Tetlock, 2002). People behave like intuitive scientists when they seek accurate explanations and understandings of true relationships in the world. Citizens as intuitive scientists is a basic assumption of democratic governance (Bimber, 2003). If citizens don't prefer facts to false rumor and myth, it can be difficult to justify the aggregation of their voices as a means of guiding government policy (Sunstein, 2006).

That human beings do not consistently behave like intuitive scientists does not mean that they never do so (Tetlock, 2002). Tetlock argues that individuals' preference for intuitive scientific behavior is *socially contingent*. That is, people behave like intuitive scientists when their social contexts provide incentives for scientific behavior. These incentives are present in some contexts and absent in others, influencing the individual's reasoning style. Of particular relevance to the study of political misinformation is the motivation to behave as *intuitive politicians* (Tetlock, 2002). People behave like intuitive politicians when they seek to maintain a positive reputation or fulfill the social duties for which they are accountable.

Tetlock argues that, far from being restricted to professional politicians, intuitive political behavior is rational and widespread. Intuitive political reasoning is also not less sophisticated. Kahan (2012) finds evidence that those who are better at sophisticated analytical reasoning are often better at defending their own social group's point of view, possibly enhancing their reputation or fulfilling an important role as a result.

Intuitive political reasoning is also not universal but is, instead, activated by social circumstances. For example, as demonstrated in the famous Asch experiments (Asch, 1951, 1956), in situations where there are strong pressures to conform, individuals willingly adopt clearly false ideas (in that case, agreeing that a shorter line is equal to a longer line) when surrounded by group members who already hold them. Individuals also modify their performance on other "objective" tasks based on information about who is going to be judging them (Lerner & Tetlock, 1999). The motivations to conform stem from the importance of maintaining a good reputation and ties to friends or social group.

Tetlock's argument raises the possibility that a key to promoting the preference for truth over falsity is the promotion of social conditions that either encourage scientific thinking and discourage political thinking or at least align the two, encouraging individuals to see it as part of their expected social role to think scientifically. Yet previous work on the correction of false rumors has not explicitly examined variation in the social context in which corrections take place. In particular, laboratory studies emphasize *asocial* corrections, such as from professional news media (e.g., Nyhan & Reifler, 2010), while computational field studies (e.g., Friggeri et al., 2014; Shin et al., 2016) have focused on the macro-social consequences of corrections, such as how much they influence the diffusion of false rumors across a social network, rather than the consequences of social context for the individual being corrected. In this article we test for the influence of one broad set of social conditions: the social network relationship between an individual who believes false information and the individual who provides them with corrective truth. Our test takes advantage of two unique affordances of social media. First, these new technologies offer the potential for a broader dissemination of facts to counter false claims (Garrett, 2011). In today's digital media environment, individuals can draw on a wealth of well-sourced and widely recognized information providers including government websites, Wikipedia, and other traditional and nontraditional sources to engage in immediate *fact-checking interventions*, in which one individual challenges the false claim of another through reference to a verified source of high-quality information (Hannak et al., 2014). Second, both the conversational dynamics and social network relationships between individuals on social media are publicly observable. This enables us to observe a set of uniquely relevant, although rare, events in which individuals engage in fact-checking within their conversations and compare outcomes based on their relationship to one another.

In this study, we collect a set of fact-checking interventions, which we refer to as "snopes," that respond to political rumors on Twitter. We then distill our observations to the subset of cases where (a) the snope is clearly a correction of a false idea; (b) the correction was unsolicited; and (c) the individual who is being corrected (the "snopee") replies to the individual who corrected them (the "snoper"), allowing us to analyze their response to the correction. We then test hypotheses about how the relationship between snoper and snopee should influence these replies based on social network theory (Burt, 2005; Coleman, 1988) and intergroup competition theory (Tajfel & Turner, 1979). While previous work has used similar techniques to identify false rumors and the overall effectiveness of corrections (e.g., Friggeri et al., 2014; Shin et al., 2016), our study differs in that it explicitly examines the influence of another variable—the social relationship between the snoper and snopee—on this effectiveness.

Social Motivations for Intuitive Science

Thinking scientifically benefits individuals in two basic contexts: (a) where individuals directly benefit from having accurate knowledge; and (b) where there are strategic incentives to scientific thinking, such as to appear to be scientific in the eyes of others. The first case is comprised of situations where an individual's beliefs inform decisions that have direct, unmediated consequences to them, and thus accurate beliefs, constructed through scientific thinking, are likely to lead to better personal outcomes than inaccurate ones. For example, in the Asch study case, a person who falsely perceives the length of a line might make mistakes in making household repairs, leading to further difficulties or distress.

This kind of direct consequence is rare in political contexts, however, because any single individual's beliefs and attitudes have so little influence over the actual direction of policy (Kahan, 2012). In democracies, an individual's personal beliefs only directly bear on government decisions in the rare case where his or her vote makes the difference in electoral outcomes (Christakis & Fowler, 2009). Furthermore, even in these cases, an individual's beliefs about one issue may not influence his or her vote (Enns, 2015). These factors remove individual incentives to be accurate. As Kahan (2012) writes,

> because what any ordinary individual believes about policy will not make a difference, the collective irrationality of ideologically motivated reasoning does not by itself create any reliable pressure or mechanism to induce

individuals to process information in a different, and morally and politically superior, way. (p. 42)

For example, in our data a number of individuals corrected one another's estimations of the U.S. federal budget deficit. Although these disputes involved huge sums, the fact that both U.S. fiscal policy and its outcomes are highly complex and decided through negotiations between many politicians makes it is hard to see that an individual personally overestimating (or underestimating) the federal deficit by (even as much as) hundreds of billions of dollars would cause them any direct consequences.

Although the direct consequences of false political information are hard to discern, individuals can experience negative *social* consequences of spreading false political information. Individuals who subscribe to false beliefs can draw scorn or distrust from peers and those who spread misinformation may not be trusted (Ellwardt, Steglich, & Wittek, 2012; Shapin, 1994). For example, in our data we observe numerous instances of people rebuking those who have shared false information. These rebukes are often followed by attempts to save face. One individual asserted that Obamacare required U.S. citizens to be implanted with microchips. Five minutes later, a snoper sent her a link to a snopes.com refutation of the claim and said "Do some homework." The snopee immediately replied to show they had done so: "I just did, right before you sent that LOL It wasn't on snopes when I first seen it. I checked."

These individual-level incentives make up only a part of people's motivation to participate in political conversation. An important power in political conversation is an individual's ability to influence others (Bond et al., 2012; Eveland, Morey, & Hutchens, 2011). Individuals can promote the diffusion of both true and false beliefs within their social communities, affecting collective outcomes.

At the collective level, however, it is not clear that truth should always be preferred. Social groups must balance the tension between within-group and between-group dynamics (Wilson, 2012). Within the group, accurate information should be preferred to inform collective decision making (Sunstein, 2006). But competition with other groups leads to different incentives. These incentives can be particularly strong in cases of intergroup *conflicts of interest*, in which groups compete for control of a common resource, like political power (Turner, 1975).

Much as there are advantages to voting or expressing preferences strategically, rather than honestly, in electoral contests (Austen-Smith & Feddersen, 2009), groups can gain from advancing false ideas that serve a strategic purpose, such as eliciting contributions to a public good (Burton-Chellew, Ross-Gillespie, & West, 2010) or mobilizing aggression (Cohen, Montoya, & Insko, 2006).

In particular, when a social cluster or group of individuals agree on myths they can support one another to take control of a collective decision, defeating competition from rival groups. Solidarity about an idea within the group can both strengthen commitment to it and intimidate rivals, irrespective of its truth content (Koudenburg, Postmes, & Gordijn, 2016). To return to the Asch study example, although an individual might be harmed by incorrectly perceiving the length of a line (for example, mistaking one foot to be 11 inches), *a group* of individuals that can unrelentingly *agree* that a foot is equal to 11 inches might gain an advantage in negotiating with other groups over things such as appropriate weights and measures for commercial commodities (Crease, 2011). As a result, groups impose pressures on individuals to express their agreement with the group as a means of demonstrating their identification with it (Kahan, 2012).

Social Structural Antecedents of Intuitive Science

The narrow set of conditions under which individuals have incentives to think scientifically about political information may appear to suggest that deliberative practice is a rare anomaly or aberration; however, this pessimistic interpretation overlooks the social structural conditions in which political conversation takes place (Morey, Eveland, & Hutchens, 2012). In fact, social network research suggests that social-structural incentives, sometimes referred to as the rationale of social capital (Burt, 2005), are the basis for many normatively preferred behaviors. For example, although honesty is an essential condition of effective communication in many contexts (Grice, 1975), it is largely encouraged by structural accountability to others who can control one's reputation and future access to information and resources (Uzzi, 1997). In fact, Shapin (1994) argues that the modern conception of intuitive science formed via social norms of interpersonal honesty within aristocratic social networks.

Similarly, attending to a diversity of views requires cognitive effort and often little personal gain. Individuals engage in it, however, as a consequence of their attempt to maintain valued social ties. That is, people who have friends who have diverse political views, or are connected to others with diverse views, attend to diverse political information as a result of their desire to do something else they value—talk with friends (Eveland et al., 2011; Lazer, Rubineau, Chetkovich, Katz, & Neblo, 2010; Song, 2015). The disincentives for such attention are thus outweighed by other, non-deliberative incentives that force diversity upon them. Individuals also appear more willing to tolerate disagreement and amend their views in light of it when conversing with friends (Morey et al., 2012).

In each of these cases, there exist social-structural conditions that compel adherence to a deliberative norm even when an individual would not, absent those structural conditions, value or adhere to that norm on their own. Similarly, social networks are likely to be a basis for the incentives to think scientifically about political information. In particular, both the dyadic relationship between the snoper and snopee and their relationship to their larger social network environment will have an impact on both reputational and intergroup mechanisms.

First, when individuals have a dyadic relationship, reputation and accountability are more likely to be important (Coleman, 1988; Lerner & Tetlock, 1999). A typical assumption in friendship relationships is that friends are "epistemic peers," meaning both assume they are equally competent at evaluating the truth of a claim (Elga, 2011). In these cases, refusal to accept correct information from a friend can lead to gossip suggesting the individual is obstinate or untrustworthy (Sommerfeld, Krambeck, & Milinski, 2008). Relationships with friends can also create safe spaces in which to reexamine one's beliefs (Morey et al., 2012). By contrast, when such corrections are given by strangers they are likely to bring substantially lower reputational risks. Thus, the incentive to think scientifically in order to save face and demonstrate one's competence is stronger when one is corrected by a friend as opposed to corrected by a stranger.

Individuals also have collective motivations for engaging in public conversation. Social media has created the opportunity for individuals to participate in networked publics where they can not only communicate information but actively build solidarity and shared understanding in a community (Papacharissi & de Fatima Oliveira, 2012; Varnelis, 2012). Since, as described in the previous section, the group can benefit from spreading false information (particularly about a rival), an important question is whether the group prefers intuitive scientists or intuitive politicians in a particular case.

The social relationship between the snoper and the snopee can be a signal of this preference. Individuals learn the norms of their group via admonitions from other group members (Cohen et al., 2006). As Kahan (2012) describes, the individual can retain or improve their standing in the group when they show a willingness to adopt or promote false views on its behalf. It follows that when other group members correct their false assertion, this utility is lost and the individual has less incentive to promote the false view.

Since friends are statistically likely (although not guaranteed) to be members of the same social group as the individual (Christakis & Fowler, 2009), a correction from a friend is likely to imply that the group does not accept this rumor and/or does not see strategic value in promoting it. Consistent with this argument, Garrett (2011) finds that individuals are more likely to believe false rumors when they are e-mailed to them by friends. By contrast, corrections from strangers are more likely to be from non-group members or even out-group members—those whose interests oppose the interests of the group (Tajfel & Turner, 1979). In these cases, individuals receive no signal suggesting that acceptance will have group benefits, and may in fact infer that acquiescence to the correction would do their group harm. In fact, Friggeri and colleagues' (2014) study of fact-checking interventions on Facebook, where the majority of interactions take place over existing friendships, indicates that although fact-checks are exceedingly rare relative to sharing of false rumors, they do lead to a significant increase in the deletion of false-rumor posts. Thus:

H1: Snopees are more likely to accept corrections from snopers who are friends than from snopers who are strangers.

Reputation and group loyalty are not earned and performed solely within the dyadic relationship between snoper and snopee. Conversations in social media are organized around larger social structures that can influence their interaction (Himelboim, 2010; Maireder & Schlogl, 2014). In this networked context, snoper and snopee are not only related dyadically (as friends or strangers) but also through ties with shared third parties (Monge & Contractor, 2003). The extent to which a pair of individuals share these triadic relationships is often described as the *embeddedness* of their relationship (Moody & White, 2003; Uzzi, 1997). Individuals within embedded relationships tend to face stronger reputational pressures because multiple third parties can report to one another on any misdeeds (Coleman, 1988). In Twitter, in particular, only individuals who follow both the snoper and the snopee can observe the entirety of their interaction. At the same time, the consequences of losing face are heightened within embedded networks (Burt, 2005). When others learn of the offense via gossip, the individual risks the loss of all of the embedded ties.

Similarly, the probability that two individuals who share many embedded ties are part of the same social group is even greater than if they simply share an existing, dyadic relationship (Monge & Contractor, 2003). Thus, snopees who are in the same embedded community as their snoper should be more likely to take the snoper's rebuke as a signal that their shared collective prefers truth in this case. For these reasons we predict the following:

H2: The larger the shared audience between snoper and snopee, the more likely that the snopee will accept the correction.

Another feature of social structure that may influence individuals' decisions to think scientifically is their social role (González-Bailón, Wang, & Borge-Holthoefer, 2014). The two-step flow model suggests most individuals consult a small number of local “experts” for political knowledge (Katz, 1957), a behavior observed in Twitter and other online forums (Choi, 2014; Himelboim, 2010; Lin, Keegan, Margolin, & Lazer, 2014). Whether these leaders or “stars” in the network are more or less likely to acquiesce to snopes than typical individuals, or “followers,” is difficult to predict a priori. On the one hand, leaders have more freedom and power to challenge other individuals. On the other hand, they are often expected to demonstrate a greater commitment to the group and uphold group norms and values (Reicher, Haslam, & Hopkins, 2005). For example, Lin and colleagues (2013) find that elites within political parties are more likely to deny political information that is obvious to the general public (that their preferred candidate performed poorly in a debate). More broadly, research indicates that the overall impact of experts on aggregate diffusion patterns is substantially influenced by the overall social structure (Watts & Dodds, 2007). Thus, there is not likely to be an easily observable aggregate signal of expert influence on misinformation, particularly since corrections are so rare (Friggeri et al., 2014). Testing the influence of potential opinion leaders thus requires examining specific cases.

In sum, while we have a clear expectation that the size of an individual’s audience—the number of followers they have—may be an important variable in determining their response to a correction, we do not have a specific prediction about the direction of this relationship. We thus propose the following research questions (RQs):

RQ1a: How does the size of the snopee’s audience affect the tendency to accept the correction?

RQ1b: How does the size of the snoper’s audience affect the snopee’s tendency to accept the correction?

Method—study 1

Data and Measures

Our procedure is similar to that used in Hannak et al. (2014). We first identify political snopes: reply tweets that have a high probability of containing corrections about a political fact. We then gather social network data—friend and follower lists—of the user who sent the reply and the user who was replied to and characterize their relationship to each other. Next we examine the dynamics of their exchange. First, we take the subset of replies in which the snopee responds. Next we code each case (that drew a response) to determine whether it is a correction. Finally, we analyze the responses to these correcting snopes to see whether the correction was accepted. These procedures are described in more detail below.

Identifying Snopes. We begin by selecting all tweets posted between January 2012 and April 2014 that meet two criteria: they (a) reply to another user, and (b) contain a reference to one of three fact-checking website domains (Snopes.com, FactCheck.org, PolitiFact.com).[1] Using these sites as indicators of corrections is common practice in fact-checking research (see Friggeri et al. [2014] and Shin et al. [2016]). We choose references to these websites in our study because, first, we expect that a high proportion of the tweets that use them in replies will be factual corrections (as compared with all replies to any rumor tweet that links to outside sources, such as newspaper articles or Wikipedia). This enables us to focus our more detailed, human coding efforts on relevant cases. Second, and importantly for our study, this method provides the benefit that all such replies contain roughly equivalent factual evidence—the standard format used on the fact-checking website. Thus, variations in acceptance of the correction should not be influenced by systematic differences in the absence/presence or quality of evidence provided.

The aforementioned studies that examine references to fact-checking websites focus on the behavior of anyone exposed to the rumor or the correction; our focus is on the specific behavior of the snopee. Thus, our method only begins with these snopes as candidates for analyzing the dynamics of correction, and drills deeper to identify a smaller set of cases where a fact-check is known to have occurred and where the corrected individual—the snopee—responds in some way.

Nomenclature—Snope Triplets. To facilitate interpretation of these conversational responses, which involve potentially confusing concepts like "replies to replies," we define our data in terms of *snope triplets*. Snope triplets are defined by three tweets:

> Tweet A0 = the "snoped" tweet—the tweet sent by the snopee that drew the snoping reply from the snoper.
>
> Tweet B1 = the "snoping" tweet—the tweet sent in reply to A0 that contains the snope.
>
> Tweet A2 = the reply from the snopee to the snoping tweet.

Identifying Political Snopes. We filter the snopes (B1s) based on the specific URLs in the replying tweets to identify those snopes that relate to U.S. politics. Specifically, for the domains Snopes.com and FactCheck.org we used all rumors within the subcategory "politics" as well as anything that had "Obama" or "Romney" in the URL. While the latter became less relevant after the 2012 election, results from Friggeri and colleagues (2014) indicate that rumors often have a long life. We use this criterion to ensure that any uncategorized rumors that included either presidential candidate were available for our analysis. For PolitiFact.com we used all tweets that contained this domain in the URL since the site is related to politics.

Characterizing Snoper-Snopee Dyads. We obtain the accounts followed and the accounts who follow each snoper and snopee in our data.[2] This allows us to characterize each relationship in terms of the following variables:

- *Mutual Friends*. When both the snoper and the snopee mutually follow each other, we characterize this relationship as "mutual friends" or, in common parlance, "friends."
- *Strangers*. When neither the snoper nor the snopee follow each other, we characterize this relationship as "strangers."
- *Snoper Follows Snopee*. If the snoper follows the snopee but the snopee does not reciprocate, we characterize this as "snoper follows snopee only."
- *Snopee Follows Snoper*. If the snopee follows the snoper but the snoper does not reciprocate, we characterize this as "snopee follows snoper only."

Although these relationships, observed in Twitter, are likely not as strong as face-to-face relationships, we expect the logic of accountability for reputation and group roles to apply, as research indicates that people are broadly aware of who their Twitter followers are (Hofer & Aubert, 2013) and that they "imagine" and adjust their tweets to satisfy their follower community when composing their tweets (Marwick & Boyd, 2011).

- *Snopee Follower Count*. The total number of Twitter followers for the snopee.
- *Snoper Follower Count*. The total number of Twitter followers for the snoper.
- *Shared Followers*. Snopers and snopees can also be characterized in terms of the number of followers they have in common. For each dyad we count the number of shared followers as the intersection of the sets of accounts that follow both the snoper and the snopee. This quantity measures the extent to which these individuals share an audience.

We note here that that the Twitter application programming interface (API) does not allow us to identify the precise timing of these friend/following relationships. That is, it is possible that after an individual fails to accept a snope from a friend, they are unfriended and become strangers, or that after an individual successfully snopes a stranger, they become friends. Recent studies on "unfriending" on Twitter suggest that this behavior is not particularly common. For example, Moon (2011) reports that unfriending occurs on just over 10% of relationships over a 50-day period. More importantly, these relationships are disproportionately newer, indicating weaker ties that, theoretically, are more like "stranger" relationships, and much less likely to occur in mutually following relationships, a finding corroborated by Xu, Huang, Kwak, and Contractor, 2013). In other work, Garimella et al (2014) find that prior to and after an "unfollow" due to a relationship dissolution the number of messages to the unfollowed party decreases.

To check for support of our assumption, we analyzed the full conversation history between each snoper and snopee in our dyad—that is, all of the times they @mentioned each other before or after the snope. Analysis revealed that conversations were robust to snoping events, as dyads that tended to converse before one snoped the other also tended to converse after the snope for both friends and strangers and irrespective of the outcome of the snope incident (β = .390, $p < .001$). This provides indirect evidence that the relationship between the snoper and snopee did not materially change with the snope event.

Identifying A0-B1-A2 Triplets. After gathering social network data we have 2,044 original snopes for which relevant social network data were available for both the snoper and the snopee. Of these, 542 contained replies from snopees (A2s; after removing 15 cases where individual sent the same reply twice). The fact that many snopes do not generate replies poses the question of how these should be treated. From one point of view non-replies might be coded as non-acceptances since the snopee provides no observable acknowledgment of the validity of the correction. Treating non-replies as non-acceptances can be problematic, however, because they outnumber replies by almost 3 to 1. This means that any systematic pattern that explains whether or not snopees reply will be likely to statistically dominate snopee tendencies within replies. In addition, Hannak et al (2014) have already studied differential rates of reply to snopes and found that, consistent with the theoretical arguments in this article, strangers are more likely to ignore (fail to reply to) snopes than friends. We observe similar dynamics here as strangers reply to fewer (24%, 266 of the 1,102) of the snopes they received compared to mutual friends (44%, 191 of the 435 snopes). Similarly, rates of reply when the snopee follows the snoper (but is not reciprocated) are similar to that of mutual friends (45%, 32 of the 71 snopes). By contrast, however, when the snoper follows the snopee (but is not reciprocated), rates of reply are quite low (13%, 53 out of 421 snopes).

Overall these figures suggest that for our primary contrast of interest—friends versus strangers—friends are not likely to be affected by differences in the rate of reply. In particular, both prior work and our data indicate that friends are more likely to respond to snopes than strangers, making an additional tendency to reply in a more accepting way further consistent with the theory. We thus exclude non-replies and focus only on the 542 cases where the snopee replied to the snope (A2 exists). Nonetheless, as a check of robustness on significant findings in our main model we also run models in which non-replies are included as non-acceptances.

Of these 542 there were 91 cases where one or more of the tweets had been deleted or was served from a protected account, or was not in English, and thus we did not code these or include them in any analysis. This left 451 complete triplets for study.

Categorizing Snopes and Responses. As Shin and colleagues (2016) report, and consistent with Hannak et al. (2014), many references to fact-checking are not corrections but promote a rumor or are responses to queries about its truthfulness (i.e., indicating the snopee is already thinking like an intuitive scientist). Our theory does not apply to these cases so they must be removed for our analysis. We code each of these 451 triplets for whether the B1 is a correction, and, if it is a correction, we code the exchange for whether the A2 accepts the correction or not. Categories are described below and additional examples provided in Table 1.

- *Non-correcting snope*. A correction points out that the snopee is wrong about a fact or spreading false information. Non-corrections took three basic forms. First, sometimes the snopee explicitly solicited facts on a topic. For example, one individual had encountered an incriminating photo of President Barack Obama and asked, "Have you seen this and know when it was taken??" Second, there were cases where the snoping tweet referred to information that was irrelevant to or actually supported, rather than corrected, the original snopee tweet. Also, since our algorithm detects any replies that refer to a snoping website, as an artifact it detects all replies to the Twitter account of snoping websites themselves, and we code these as "non-correcting" since they do not actually cite fact-checking corrections.
- *Accepts*. The snopee explicitly acknowledges that the fact presented by the snopee is correct or clearly accepts it as correct by, say, admitting that they were wrong. Alternatively, the snopee accepts the legitimacy of the fact as possibly true, stating that it is plausible or worthy of consideration. This case thus captures whether the snope was "successful" insofar as it modified the snopee's sense of what is true, or at least of what it is acceptable to claim as true. Importantly, the snopee need not be convinced of the validity of a larger argument or new political position; they must simply acknowledge that the fact in the snope is accurate (see Figure 1 and Table 1 for examples).
- *Ambiguous or Other*. This includes cases where the snopee rejects the correction explicitly as well as other cases that were difficult to classify in epistemic terms. For example, sometimes the snopee does not comment on the truth or falseness of the snope fact at all but shows hostility toward the correction or snoper. This includes statements like "I don't care if it's true" as well as statements that personally attack the snoper's intentions or change the subject without making a statement about the fact itself. This also includes cases where it is too difficult to interpret what the snopee is saying, often because they use slang or refer to entities or ideas unfamiliar to the coder. We initially attempted to code these cases into subcategories of

Table 1
Coding scheme

Category	Typical Forms	Example
Non-correcting snope	Original (A0) tweet solicits fact-check or further evidence for a claim	"@mention sounds too funny to be true"
	Snope (B1) corroborates original (A0) tweet	Snoper: "@snopee even more appalling. How this didnt make national news in 2007 is beyond me http://url"
	Replies to fact-checking website	"@factcheckdotorg Hey the link is broken, may want to repost with working link. http://url"
Accepts	Reply (A2) accepts snope (B1) as true or correct	"@snoper Sorry I was incorrect, but i would argue that march of 2008, 4 years ago, gas prices comparable with a steep decline" [sic]
	Reply (A2) considers (B1) is plausible	"@snoper Snopes is owned by George Soros. However, it can be debated that the ring says … No God but God instead"
Ambiguous or other	Reply (A2) rejects snope (B1)	"@snoper Snopes is as left wing oriented as they come & thus their 'facts' tend to reflect their politics"
	Reply (A2) is ambiguous with respect to truth of snope (B1)	"@snoper The success that Romney had had in his life is the reason you hate him. How come your link didn't cover executive orders?"

"rejection" of the snope and "other." However, obtaining a consistent coding protocol was difficult because "rejection" of a fact, the truth of a claim, was rare in a pure form and hard to meaningfully distinguish from the "rejection" of the social behavior being corrected. For example, it is unclear if the individual in the top left corner of Figure 1 ("all I wanna say is") is rejecting the correction or simply ignoring or moving on from it. Similarly, the tweet in the last row of Table 1 ("how come your link didn't") argues with the snoper but doesn't quite indicate rejection of the truth of the article. Since our research question focuses on the positive outcome (acceptance), we did not pursue coding rejections further.

Coding was conducted by two coders (the author and a research assistant). The coders independently evaluated a subset of tweets (about 10%) and met to discuss and recode them until they were in agreement regarding the coding protocol. Then the remaining snope triplets were coded independently. The coders agreed on 71% of cases across the three mutually exclusive categories (non-correcting, accepts, ambiguous or other), yielding a Cohen's kappa = .58. Since this level shows only "moderate" agreement it was decided that further analysis would only be conducted on the cases where both coders

Figure 1. Frequency of Acceptance by Social Relationship

agreed with the characterization of the snope (correcting, non-correcting) and the snopee's acceptance of the fact. This yields 322 triplets, of which 229 (71%) are corrections and thus suitable for analysis of whether a snopee "accepts" a fact. As a check on robustness, we also analyze the full (451) cases from the point of view of each coder's interpretation (whether the snope was a "correction" and whether it was "accepted") and obtain substantively similar results (see supplemental Appendix).

Analysis

Statistical tests are conducted using logistic regression. The dependent variable is whether or not each snopee's response to the snope is an acceptance. The relationships between snoper and snopee are treated as covariates in this analysis. We also run ordinary least squares (OLS) regression and find substantively similar results (see supplemental Appendix). Although the dependent variable is binary, there is now evidence that OLS provides appropriate results in most cases (Angrist & Pischke, 2009). We use the OLS regression to provide more easily interpretable information about the impact of our theorized variables on the absolute level of correction acceptance, rather than just their impact on relative odds.

Results—study 1

Preliminary Analysis

As previously described, our coding system yielded 322 cases where there was complete data and agreement about whether the snope was a correction. Of these, 93 cases were "non-corrections." Although this is a substantial loss in terms of statistical power, it did not appear to introduce any bias, as the rate at which corrections and non-corrections come from friends and strangers is remarkably similar, with 70.4% of snopes from friends being corrections and 72.5% of snopes from strangers being corrections (chi-square = .60, $df = 1$, $p = .56$). Our remaining analysis focuses on the 229 corrections within this set. In these 229 cases, 135 (59%) were instances where the snopee "accepted" the snope as true.

Hypothesis Tests

Hypothesis 1. Hypothesis 1 (H1) predicted that snopees would be more willing to accept when snopes came from friends as opposed to from strangers. We begin our analysis by examining the impact of the dyadic relationship between snoper and snopee. Figure 1 shows a breakdown of whether the snopee accepts the snope fact according to whether the snopee and snoper had one of two relationships: snopee and snoper each follow each other (mutual friends), and neither follows each other (strangers). We see that for stranger-to-stranger snopes, the percentage in which the fact is accepted is 39%, whereas it is more than 73% for cases where snopes are between mutual friends. This trend is similar for the other relationship types (not shown in the figure for visual clarity): 7 out of 11 (64%) of snopees who follow their snopers accept the fact, and 23 out of 29 (79%) of snopees who are followed by their snopers accept the fact.

Regression results highlighting these variables, shown in column 1 of Table 2, indicate that when the snoper follows the snopee, either as a mutual friend or as an unreciprocated follower, this significantly increases the likelihood of the snopee accepting a snoper's fact as true. Although they are somewhat reduced, these results are robust when controlling for other hypothesized variables (Table 2, column 4), with the odds of a mutual friend accepting a correction being 2.5 times ($\beta = .899$, $p < .05$) that of a stranger accepting a correction. In terms of absolute rates, OLS regression results indicate that strangers have a 24% chance of accepting a correction, with mutual friends increasing the chance by 22% to 46%. This result holds in each of our robustness checks—when non-responses are treated as "non-acceptances" and when different coders' perspectives are accounted for (see supplemental Appendix). We also observe that when the snoper follows the snopee exclusively the odds of the snopee accepting the correction are 3.4 times greater ($\beta = 1.211$, $p < .05$). However, these results do not hold when non-responses are treated as "non-acceptances," possibly due to the fact that these particular relationships garner fewer replies (only 13% as reported earlier). It is thus difficult to rule out the possibility of selection effects here, with snopees only replying to their (unreciprocated) followers when they plan to agree. Nonetheless, the robust support for mutual friends provides support for H1.

Hypothesis 2. Hypothesis 2 predicted that snopees would respond more productively when they were embedded in a shared audience with snopers. We measure this shared audience by counting the number of Twitter followers shared by the snoper and snopee. Results for this variable on the acceptance of the snope fact are shown in column 2 of Table 2. There is a statistically significant relationship ($\beta = .449$, $p < .001$). The larger the number of followers in common with the snoper, the more likely the snopee will accept the fact in the snope. These bivariate results hold in each of our robustness tests. The number of shared followers is also significant in the full model (Table 2, column 4) ($\beta = .265$, $p < .01$); however, these results are not significant in any of the robustness tests (see supplemental Appendix). Thus, overall these results indicate partial support for H2.

Table 2
Study 1—Do snopees accept the snoping fact as true?

	Dependent Variable:			
	Accepts			
	(1)	(2)	(3)	(4)
Mutual friends	1.450***			0.899*
	(0.315)			(0.415)
Snopee follows snoper	1.009			0.742
	(0.661)			(0.680)
Snoper follows snopee	1.639***			1.211*
	(0.480)			(0.508)
Shared followers (log)		0.449***		0.265*
		(0.095)		(0.131)
Snopee follower count (log)			0.194*	0.112
			(0.089)	(0.101)
Snoper followers (log)			0.054	–0.022
			(0.086)	(0.098)
Constant	–0.450*	–0.366	–1.148	–1.172
	(0.210)	(0.199)	(0.703)	(0.834)
Observations	229	229	229	229
Log likelihood	–141.150	–141.959	–152.189	–137.108
Akaike information criterion	290.301	287.919	310.377	288.215

Note. Logistic regression.
* $p < 0.05$.*** $p < 0.001$.

Research Question 1. Research question (RQ) 1a asked whether the size of a snopee's audience had an influence on the snopee's decision to accept the snope fact and/or to defend their original statement. RQ1b asked whether the size of a snoper's audience had an influence on these decisions by the snopee. Column 3 of Table 2 shows a significant bivariate relationship between the number of followers possessed by the snopee and their likelihood of accepting the snope fact ($\beta = .194$, $p < .05$) and is reversed (i.e., negative) in our robustness checks using all replies, possibly due to snopees with large followings being unlikely to send replies. The relationship is also no longer significant when accounting for the other hypothesized variables (column 4). Column 3 also shows no statistically significant relationship between the snoper's audience size and the likelihood that the snopee will accept the fact.

Method—study 2

The results of Study 1 were limited in two important respects. First, all of the data were collected from one particular time period that featured an electoral competition between Barack Obama and Mitt Romney. It is possible that observed differences are due to the kinds of rumors and corrections that related to those candidates, rather than a broader principle. Second, by restricting our data collection to only politically relevant tweets, we were not able to compare political snoping to nonpolitical snoping. In particular, while both kinds of snopes should be subject to reputational mechanisms, intergroup competitive dynamics should be more salient for political snopes where there is a clear conflict of interest between parties (Turner, 1975), and thus comparing political to nonpolitical snopes should shed light on the relative importance of these mechanisms.

We thus attempted to replicate our significant findings on a new data set. On February 3, 2016, we manually scraped the Twitter website for tweets that contained hyperlink references to the fact-checking site Snopes.com. Manual scraping enables quick access to recent data but does not afford access to longer term archives like Firehose does. Since the former was sufficient for our purposes, we selected this method.

Data going back to October 31, 2015, was obtained by repeatedly scrolling down for more search results in the Web interface and using Chrome Developer Tools to save the Data Object Model. We then subset these tweets to include only those that contain our snope triplet structure in which one tweet is snoped and then the snopee replies to this snoping tweet.

As stated earlier, we loosened our filtering criteria to include both political and nonpolitical snopes. Snopes were categorized as political if they referred to rumors marked as "political" on Snopes.com or if they contained reference to a known politician such as Barack Obama, Hillary Clinton, or Donald Trump.

To further check for the robustness of our methods, these tweets were then coded by workers for Crowdflower.com, rather than the original set of coders. CrowdFlower is a crowdsourcing platform on which paying clients can submit micro tasks, also referred to as Human Intelligence Tasks (HIT), and paid contributors/coders can sign up to work on these tasks for pay. Each task is typically a small set of multiple-choice questions.

Our scraper returned 759 tweets with the triplet structure. Of these, 465 of the snoping tweets were identified consistently by two coders as an unsolicited correction to a rumor. We then deployed four or five coders to code each of these 465 for whether the snopee accepted the correction. Average agreement was 91%. Since the coding was performed by many different individuals (23 different individuals participated on at least one snoping case), we calculate Krippendorff's alpha = .661 (Krippendorff, 1980). We then selected only the 414 triplets where at least 75% of the coders agreed on the category for further analysis (i.e., 3 out of 4, or 4 out of 5).

Results—study 2

To examine the robustness and generalizability of our finding from Study 1 we retest H1 and H2 using the same model specification on the Study 2 data. Results, shown in Table 3, are substantively similar to those found in our original data set. As in our original test of H1, the coefficient for mutual friends is significant in all models, with mutual friends increasing the odds of acceptance by 5.1 times over strangers in the full model comparable to that in Study 1 (column 4) ($\beta = 1.633$, $p < .001$). Comparison of the OLS coefficients indicates that the magnitude of the change is also similar. Mutual friends are estimated to increase the probability of acceptance by 22% in Study 1 and 31% in Study 2 (see supplemental Appendix). We also observe a statistically significant coefficient when the snoper follows the snopee exclusively. However, the robustness analysis from Study 1 suggests this is likely subject to selection bias and we do not interpret further.

Also, as in our original test of H2, the number of common followers also showed a significant, positive relationship to acceptance when considered on its own (Table 3,

Table 3
Study 2—Do snopees accept the snoping fact as true?

	Dependent Variable:					
	Accepts					
	(1)	(2)	(3)	(4)	(5)	(6)
Mutual friends	1.610***			1.633***	1.496***	1.388**
	(0.260)			(0.311)	(0.319)	(0.504)
Snopee follows snoper	0.759			0.675	0.602	0.597
	(0.608)			(0.632)	(0.639)	(0.641)
Snoper follows snopee	1.922***			2.108***	2.030***	2.019***
	(0.407)			(0.441)	(0.445)	(0.447)
Shared followers (log)		0.152**		0.024	0.018	0.020
		(0.055)		(0.088)	(0.089)	(0.089)
Snopee follower count (log)			–0.041	–0.151	–0.137	–0.137
			(0.066)	(0.083)	(0.084)	(0.084)
Snoper followers (log)			0.003	–0.030	–0.019	–0.023
			(0.062)	(0.080)	(0.080)	(0.081)
Political rumor					–0.558*	–0.639
					(0.270)	(0.396)
Mutual friends x political rumor						0.150
						(0.544)
Constant	–1.857***	–1.407***	–0.764	–0.715	–0.398	–0.311
	(0.190)	(0.184)	(0.591)	(0.703)	(0.727)	(0.792)
Observations	414	414	414	414	414	414
Log likelihood	–213.400	–234.908	–238.466	–211.410	–209.313	–209.275
Akaike information criterion	434.801	473.817	482.932	436.821	434.625	436.550

Note. Logistic regression.
* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

column 2; $\beta = .152$, $p < .01$), but the effect was very small, increasing the probability of acceptance by only 3% in the OLS model. Also, this relationship was not significant in the full model. Thus, H2, which received weak support in Study 1, does not obtain support in Study 2, suggesting that this mechanism is not independent of that which produces H1.

Post Hoc Analyses

Attitude Consistency. One possible explanation for our findings is that snopes from friends are more likely to be accepted because friends are more likely than strangers to correct rumors that the snopee does not wish to be true. If this is correct, then existing social relationships may still be important to fact-checking interventions, but not necessarily because they encourage the snopee to think more scientifically.

Because we perform our analysis only on cases where the rumor is asserted and exclude cases where a snope is "solicited" (via the expression of doubt or a request for corroboration), we expect most rumors to be attitude consistent, and thus most corrections to be resisted regardless of the source. Nonetheless, to explore this possibility we examine the extent to which the rumors shared by the snopee are consistent with their attitudes.

Measuring the attitude consistency of a rumor is difficult with our data because (a) we do not have access to attitude surveys for our subjects, and so any categorization is speculative; and (b) inspection of specific rumors shows that many do not align clearly with easily identified attitudes such as partisanship. For example, the third most-shared rumor in our data is a fake photo with Megyn Kelly (formerly of Fox News) and a man who is purported to be a Saudi prince. It is not clear what prior information available on an individual's Twitter account would indicate whether this rumor is consistent with their attitudes.

We nonetheless categorize a modest subset of our rumors where attitude consistency is easier to define in order to get a descriptive sense of attitude constancy patterns in our data. We examine all cases of rumors and snopes related to either party's leading presidential candidate—Donald Trump and Hillary Clinton (82 total, for which 74 had profile information available). We then look for the declaration of a partisan identity (e.g., "Republican") or a candidate preference ("Trump supporter") on the snopee's profile page. Coding was performed independently by two of the authors (Cohen's kappa = .83 after first round of coding, with further clarification of criteria leading to 100% agreement). In our data there are 30 snopes with reference to a rumor about Trump and, of these, 16 were attitude consistent (meaning the rumor was negative about Trump and was sent by a Democrat or was positive about Trump and sent by a Republican or Trump supporter), and 0 being attitude inconsistent. There were also 14 uncategorizable based on the user's profile information (i.e., they made no mentions of candidates, parties, or political loyalites on their profile). Similarly, there are 44 snopes with reference to a rumor about Clinton and, of these, 32 were attitude consistent and 0 were attitude inconsistent, with 12 being uncategorizable. Although the uncategorized profiles make it difficult to infer the extent to which all rumors are attitude consistent, the fact that we find no evidence that any rumors are attitude *in*consistent suggests that variations in attitude consistency are not likely to explain the entirety of our results.

A second test is to consider only these snopes of attitude-consistent rumors to see if we recover similar patterns as in our full data set. Similar to our main result, we find that only 13% of snopes of attitude-consistent rumors from strangers are accepted compared with 40% of snopes of attitude-consistent rumors from non-strangers (chi-square = 3, $df = 1$, $p = .08$). If we restrict the non-strangers to mutual friends the percentage acceptance increases to 60% (chi-square = 6.86, $df = 1$, $p < .01$). From these analyses we conclude that while variations in the attitude consistency of rumors snoped across relationship types may explain a portion of our results, they are not likely to explain them entirely. Nonetheless, as we acknowledge in the limitations, this question requires more thorough study.

Comparing Reputational Versus Group-Related Motivations. As we argued in the theoretical section, both H1 and H2 can be justified by two distinct mechanisms: the snopee's concern for their reputation and the snopee's concern for their group's ability to compete with others. Study 2 enables us to compare the influence of reputational mechanisms, which should apply equally to both kinds of rumors, and group competition-based mechanisms, which should be stronger for political rumors. Specifically, retaining false ideas and rejecting truth should be a blemish to a snopee's reputation whether or not the information is political. By contrast, the motivation to accept corrections from friends, or to reject corrections from strangers, out of a sense of group morality (Cohen et al., 2006) should be stronger for political rumors, particularly during an election season when groups are competing for resources and ideological control (Kahan, 2012).

As described earlier, we define political rumors as those for which the snopes URL includes "politics" or the name of an elected official or candidate. Of our 414 cases, 317 were political and 97 were not. We first observe that stranger-stranger snopes outnumber friend-friend snopes by more than 2 to 1 for political corrections (205 to 78), yet are only a bit more than half of those for nonpolitical corrections (32 to 52; $\chi^2 = 31.91$, $df = 1$, $p < .001$). Results also indicate that, overall, the rate of acceptance of political rumors is lower (21%) than the rate of acceptance of non-political rumors (42%, $\chi^2 = 15.53$, $df = 1$, $p < .001$).

Further analysis indicates that the snoper-snopee relationship and the nature of the topic of the correction have an independent relationship to acceptance, although not a significant interaction (see Table 3, column 6). Column 5 of Table 3 shows the results when "political" is included as a predictor of acceptance in the main model. Mutual friends remains significant, while the political variable also shows a significant, negative effect ($\beta = -.558$, $p < .05$; odds ratio = .57). Examination of the OLS models shows that the magnitude of the coefficients for snoper-snopee relationship and the nature of the correction (comparable for these dummy variables) suggests that the reputational force of accountability is stronger. Specifically, a political rumor correction from a friend was more likely to be accepted than an apolitical correction from a stranger (27.6% increase for friendship – 11.2% decrease for the rumor being political = 16.4% increase in probability of acceptance). Again, we have no evidence of a significant interaction, however, providing no evidence that the relationship matters even more when corrections are political, as implied by the intergroup competition explanation.

Discussion

Review of Findings

Our study departs from recent work on rumors and fact-checking by examining the behavior of the corrected person as opposed to those in the social network as a whole (e.g., Friggeri et al., 2014; Shin et al., 2016). In this study we argued that since individuals feel more accountable to their friends and shared community, and are more likely to share collective interests with them, they should be more likely to respond "scientifically"—

accepting facts that challenge their statements—when they have a bond or share a community with the person who has corrected them. The results support these claims. Across two different data sets collected more than three years apart, we find support for the hypothesis that when the people involved in the correction have a mutual relationship, the correction is more likely to be accepted. Controlling for this direct relationship, however, there was only limited evidence that the extent to which the individual responded to the correction in front of a common community or audience had an effect. There was also little evidence that the size of a snoper's or snopee's audience played a significant part in the fact-checking response.

We also compared the responses to corrections of political rumors to corrections of nonpolitical rumors. Results indicate that while political rumor corrections are less likely to be accepted overall, the dynamics for both kinds of corrections are similar: corrections from friends and followers are more likely to be accepted. In particular, a political correction from a friend was more likely to be accepted than an apolitical correction from a stranger.

Limitations

As described in the method section, the Twitter API does not provide information on when snoper-snopee friendships are established or disbanded. Our analysis revealed indirect evidence that their relationships did not substantially change as the result of the snoping incident; however, in future work it might be of interest to apply existing approaches to estimate the date when a social link was established (Meeder et al., 2011). Controlled experiments in which existing friends and strangers are recruited to correct one another could also eliminate this confound.

As noted earlier, our method cannot completely rule out the possibility that friends are correcting friends in a different manner from how strangers are correcting strangers, either because their corrections are attitude consistent or on less controversial rumors. These effects would still be consistent with the idea that fact-checking is socially contingent; however, the theoretical mechanism would be somewhat different. Specifically, instead of activating intuitive science, friends may be sensitive to one another's "latitude of acceptance" on emotionally charged topics (Sherif & Hovland, 1961). While we provide some evidence that our hypotheses hold even when controlling for attitude consistency, a proper test would require true prior information about the snopee's attitude in relation to the rumor before the correction. These kinds of data would be more appropriately obtained in a controlled experiment design where attitudinal pretests could be given and then corrections issued from friends or strangers via experimental manipulation.

Another limitation is that we focus on the dyadic, "local" relationship between the snoper and snopee. Although we also include information about the common audience size in our models, more "global" network information could also be included to quantify their closeness, such as their path-distance to one another. Similarly, one could try and infer a Twitter user's political preference and look at whether two users are in the same political camp, regardless of their dyadic relationship (Conover et al., 2012) This approach brings its own limitation, as it is not clear that individuals who do not know one another would be aware of one another's political loyalties. Nonetheless, this is a direction worth pursuing further.

Last but not least our results are obtained using one particular online social network, Twitter, in a specific political context—U.S. presidential politics—and so the question of generalizability naturally arises. To address this we have embedded our work in the wider frame of research on rumors and fact-checking. As our findings are consistent with this body of work, we believe that our findings are sound and not a domain or data-source-induced artifact. Nonetheless, future work should consider testing these ideas both in new platforms, including, if possible, face-to-face rumor corrections, and well as new political contexts, such as those in other countries or outside of the context of a hotly contested election.

Broader Implications

Our results suggest that, consistent with Tetlock's (2002) argument about social contingency, people's decisions to behave like intuitive scientists when engaging with political information depends substantially on the social structural context in which conversations take place. One implication of this is that the source of the ineffectiveness of corrections observed in public discourse (Shin et al., 2016) may be due to the social position of these corrections, rather than an innate tendency for people to resist facts. In other words, corrections may appear to be ineffective because corrections from strangers are both more common and less likely to be accepted. These tendencies will serve to paint a bleak picture in any data set that does not account for the snoper-snopee relationship. Moreover, many data sets (e.g., Friggeri et al., 2014) include relational data regarding snopers, snopees, and third-party observers. According to the theory we have proposed here, we should observe differences not only in the behavior of the snopee after being snoped by a friend but also in the behavior of all other friends of the snoper.

This raises the question of whether the dynamics of face-to-face rumor correction are similar to those observed on Twitter in this study. It may be that stranger-stranger corrections are dominant on Twitter because of the unique opportunity that social media provide for individuals to carry on conversations with complete strangers. In other conversational settings, people spend their time talking to their friends rather than to strangers (Eagle, Pentland, & Lazer, 2009). Thus, it may be that friend-friend corrections are more common than stranger corrections in face-to-face settings, and thus acceptance of facts is more common than social media and laboratory studies might lead us to conclude. However, it may also be the case that intergroup competition motivations deter friend-friend corrections (or encourage stranger-stranger corrections), such that friends are reluctant to correct one another, or to even investigate one another's claims, when they share group-enhancing myths. Thus testing whether the dynamics we have observed on Twitter hold in face-to-face conversations is an important area of future research.

The importance of social context in encouraging intuitive science also suggests a theoretical shift in the study of misinformation. In typical research on misinformation correction individuals are presented with corrections from "neutral" sources like researchers or professional media such as newspapers. Our results suggest that a key variable missing in these studies is accountability—a motivation to have accurate information about a topic where there is no individual consequence to having incorrect knowledge. This can be operationalized with friendship relationships, as in this study, or in other ways where accountability is manipulated (see Lerner & Tetlock, 1999).

More broadly these findings suggest that, as with recent concerns over the selective exposure to homogeneous political content (Bakshy, Messing, & Adamic, 2015; Bennett & Iyengar, 2008; Stroud, 2010), the prevalence of political misinformation may be borne of homophily in social networks. Friendship ties across ideological groups may be able to serve as a bridge, wherein people accept the validity of criticism even from those with whom they generally disagree. The fact that, in our results, the dyadic relationship between snoper and snopee had a much stronger effect than the extent to which they were embedded in a shared community suggests that relationships that bridge communities can have this effect.

Conclusion

This study tested, and found evidence for, the idea that social network relationships are important conditions for determining whether individuals will accept corrections to political misinformation using real-world conversations in which individuals corrected one another. As with similar findings on the role of networks in political discussion, it is hoped that this work will stimulate further research to better understand when and how misinformation corrections can be made most effective.

Notes

1. Using the Twitter Firehose, access to tweet IDs provided courtesy of Elad Yom-Tov of Microsoft Research.
2. Twitter's public REST API.

Acknowledgments

The authors would like to thank Elad Yom-Tov for his generosity in supplying tweet IDs, as well as Bruce Desmarais, Brian Keegan, David Lazer, and four anonymous reviewers for their helpful comments on this work.

Funding

This research was supported in part by a grant from the Cornell Institute for Social Sciences.

Supplemental Material

Supplemental data for this article can be accessed on the publisher's website at: https://doi.org/10.1080/10584609.2017.1334018.

References

Allport, G. W., & Postman, L. (1947). *The psychology of rumor*. Oxford, UK: Henry Holt.

Angrist, J. D., & Pischke, J.-S. (2009). Mostly harmless econometrics: an empiricist's companion (Vol. 1). Princeton, NJ.: Princeton university press Princeton.

Asch, S. E. (1951). Effects of group pressure on the modification and distortion of judgments. In H.Geutzkow (Ed.), *Groups, leadership, and men*. 222–236, Pittsburgh, PA: Carnegie.

Asch, S. E. (1956). Studies of independence and conformities. *Psychological Monographs*, General and Applied, *70*(9), 1–70.

Austen-Smith, D., & Feddersen, T. J. (2009). Information aggregation and communication in committees. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1518), 763–769. doi:10.1098/rstb.2008.0256

Bakshy, E., Messing, S., & Adamic, L. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132.

Bennett, W. L., & Iyengar, S. (2008). A new era of minimal effects? The changing foundations of political communication. *Journal of Communication*, *58*(4), 707–731. doi:10.1111/jcom.2008.58.issue-4

Bessi, A., Coletto, M., Davidescu, G. A., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Science vs. conspiracy: Collective narratives in the age of misinformation. *PLOS ONE*, *10*(2), e0118093. doi:10.1371/journal.pone.0118093

Bimber, B. A. (2003). *Information and American democracy: Technology in the evolution of political power*. New York, NY: Cambridge University Press.
Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*(7415), 295–298. doi:10.1038/nature11421
Burt, R. S. (2005). *Brokerage and closure: An introduction to social capital*. Oxford, UK. Oxford University Press.
Burton-Chellew, M. N., Ross-Gillespie, A., & West, S. A. (2010). Cooperation in humans: Competition between groups and proximate emotions. *Evolution and Human Behavior*, *31* (2), 104–108. doi:10.1016/j.evolhumbehav.2009.07.005
Choi, S. (2014). Flow, diversity, form, and influence of political talk in social-media-based public forums. *Human Communication Research*, *40*(2), 209–237. doi:10.1111/hcre.12023
Christakis, N. A., & Fowler, J. H. (2009). *Connected: The surprising power of our social networks and how they shape our lives*. New York/Little Brown.
Cohen, T. R., Montoya, R. M., & Insko, C. A. (2006). Group morality and intergroup relations: Cross-cultural and experimental evidence. *Personality and Social Psychology Bulletin*, *32* (11), 1559–1572. doi:10.1177/0146167206291673
Coleman, J. S. (1988). Social capital in the creation of human-capital. *American Journal of Sociology*, *94*, S95–S120. doi:10.1086/228943
Conover, M. D., Gonçalves, B., Flammini, A., & Menczer, F. (2012). Partisan asymmetries in online political activity. *EPJ Data Science*, *1*(1), 1–19. doi:10.1140/epjds6
Crease, R. P. (2011). *World in the balance: The historic quest for an absolute system of measurement*. New York/W.W. Norton.
Eagle, N., Pentland, A. S., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, *106*(36), 15274–15278. doi:10.1073/pnas.0900282106
Einwiller, S. A., & Kamins, M. A. (2008). Rumor has it: The moderating effect of identification on rumor impact and the effectiveness of rumor refutation. *Journal of Applied Social Psychology*, *38*(9), 2248–2272. doi:10.1111/jasp.2008.38.issue-9
Elga, A. (2011). Reflection and disagreement. In A. I. Goldman & D. Whitcomb (Eds.), *Social epistemology: Essential readings* (pp. 158–182). Oxford, England; New York, NY: Oxford University Press.
Ellwardt, L., Steglich, C., & Wittek, R. (2012). The co-evolution of gossip and friendship in workplace social networks. *Social Networks*, *34*(4), 623–633. doi:10.1016/j.socnet.2012.07.002
Enns, P. K. (2015). Relative policy support and coincidental representation. *Perspectives on Politics*, *13*(04), 1053–1064. doi:10.1017/S1537592715002315
Eveland, W. P., Morey, A. C., & Hutchens, M. J. (2011). Beyond deliberation: New directions for the study of informal political conversation from a communication perspective. *Journal of Communication*, *61*(6), 1082–1103. doi:10.1111/j.1460-2466.2011.01598.x
Friggeri, A., Adamic, L. A., Eckles, D., & Cheng, J. (2014). Rumor cascades. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media* (pp. 101–110). Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/download/8122/8110
Garrett, R. K. (2011). Troubling consequences of online political rumoring. *Human Communication Research*, *37*(2), 255–274. doi:10.1111/j.1468-2958.2010.01401.x
Garrett, R. K., Nisbet, E. C., & Lynch, E. K. (2013). Undermining the corrective effects of media-based political fact checking? The role of contextual cues and naïve theory. *Journal of Communication*, *63*(4), 617–637. doi:10.1111/jcom.12038
Giglietto, F., Iannelli, L., Rossi, L., & Valeriani, A. (2016). *Fakes, news and the election: A new taxonomy for the study of misleading information within the hybrid media system*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2878774
González-Bailón, S., Wang, N., & Borge-Holthoefer, J. (2014). The emergence of roles in large-scale networks of communication. *EPJ Data Science*, *3*(1), 1–16. doi:10.1140/epjds/s13688-014-0032-y

Gottfried, J. A., Hardy, B. W., Winneg, K. M., & Jamieson, K. H. (2013). Did fact checking matter in the 2012 presidential campaign?*American Behavioral Scientist*, *57*(11), 1558–1567. doi:10.1177/0002764213489012

Grice, H. P. (1975). *Logic and conversation*. In P. Cole and J.L. Morgan (Eds.), Syntax and semantics, Vol. 3: Speech acts (pp. 225–242). New York: Seminar Press.

Hannak, A., Margolin, D., Keegan, B., & Weber, I. (2014). Get back! You don't know me like that: The social mediation of fact checking interventions in Twitter conversations. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media* (pp. 187–196). Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPDFInterstitial/8115/8119

Himelboim, I. (2010). Civil society and online political discourse: The network structure of unrestricted discussions. *Communication Research*, *38*(5), 634–659. doi:10.1177/0093650210384853

Hofer, M., & Aubert, V. (2013). Perceived bridging and bonding social capital on Twitter: Differentiating between followers and followees. *Computers in Human Behavior*, *29*(6), 2134–2142. doi:10.1016/j.chb.2013.04.038

Kahan, D. M. (2012). Ideology, motivated reasoning, and cognitive reflection: An experimental study. *Judgment and Decision Making*, *8*, 407–424.

Katz, E. (1957). The two-step flow of communication: An up-to-date report on an hypothesis. *Public Opinion Quarterly*, *21*(1), 61–78. doi:10.1086/266687

Koudenburg, N., Postmes, T., & Gordijn, E. H. (2016). Beyond content of conversation: The role of conversational form in the emergence and regulation of social structure. *Personality and Social Psychology Review*, (21) 50–71. doi:10.1177/1088868315626022

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage.

Lazer, D., Rubineau, B., Chetkovich, C., Katz, N., & Neblo, M. (2010). The coevolution of networks and political attitudes. *Political Communication*, *27*(3), 248–274. doi:10.1080/10584609.2010.500187

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*(2), 255–275. doi:10.1037/0033-2909.125.2.255

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. doi:10.1177/1529100612451018

Lin, Y., Margolin, D., Keegan, B., & Lazer, D. (2013). Voices of Victory: A Computational Focus Group Framework for Tracking Opinion Shift in Real Time. *22nd International World Wide Web Conference* (pp. 737–748). Rio.

Lin, Y.-R., Keegan, B., Margolin, D., & Lazer, D. (2014). Rising tides or rising stars? Dynamics ofshared attention on Twitter during media events. *PLOS ONE*, *9*(5), e94093. doi:10.1371/journal.pone.0094093

Maireder, A., & Schlogl, S. (2014). 24 hours of an #outcry: The networked publics of a socio-political debate. *European Journal of Communication*, *29*(6), 687–702. doi:10.1177/0267323114545710

Marwick, A. E., & Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114–133. doi:10.1177/1461444810365313

Meeder, B., Karrer, B., Sayedi, A., Ravi, R., Borgs, C., & Chayes, J. (2011). We know who you followed last summer: Inferring social link creation times in Twitter. In *Proceedings of the 20th International Conference on World Wide Web* (pp. 517–526). New York, NY: ACM. doi:10.1145/1963405.1963479

Mocanu, D., Rossi, L., Zhang, Q., Karsai, M., & Quattrociocchi, W. (2015). Collective attention in the age of (mis)information. *Computers in Human Behavior*, *51*, 1198–1204. doi:10.1016/j.chb.2015.01.024

Monge, P., & Contractor, N. (2003). *Theories of communication networks*. New York, NY: Oxford University Press.

Moody, J., & White, D. R. (2003). Structural cohesion and embeddedness: A hierarchical concept of social groups. *American Sociological Review*, *68*, 103–127. doi:10.2307/3088904

Moon, S. (2011). Analysis of Twitter unfollow: How often do people unfollow in Twitter and why? In *International Conference on Social Informatics* (p. 7). Springer.

Morey, A. C., Eveland, W. P., & Hutchens, M. J. (2012). The "who" matters: Types of interpersonal relationships and avoidance of political disagreement. *Political Communication*, *29*(1), 86–103. doi:10.1080/10584609.2011.641070

Mutz, D. C. (2008). Is deliberative democracy a falsifiable theory? *Annual Review of Political Science*, *11*(1), 521–538. doi:10.1146/annurev.polisci.11.081306.070308

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330. doi:10.1007/s11109-010-9112-2

Papacharissi, Z., & de Fatima Oliveira, M. (2012). Affective news and networked publics: The rhythms of news storytelling on #Egypt. *Journal of Communication*, *62*(2), 266–282. doi:10.1111/j.1460-2466.2012.01630.x

Reicher, S., Haslam, S. A., & Hopkins, N. (2005). Social identity and the dynamics of leadership: Leaders and followers as collaborative agents in the transformation of social reality. *The Leadership Quarterly*, *16*(4), 547–568. doi:10.1016/j.leaqua.2005.06.007

Shapin, S. (1994). *A social history of truth: Civility and science in seventeenth-century England*.

Sherif, M., & Hovland, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change*, University of Chicago.

Shin, J., Driscoll, K., & Bar, F. (2016). Political rumoring on Twitter during the 2012 U.S. presidential election: Rumor diffusion and correction. *New Media & Society*, 1–22. doi:10.1177/1461444816634054

Sommerfeld, R. D., Krambeck, H.-J., & Milinski, M. (2008). Multiple gossip statements and their effect on reputation and trustworthiness. In *Proceedings of the Royal Society B: Biological Sciences*, *275*(1650), 2529–2536. doi:10.1098/rspb.2008.0762

Song, H. (2015). Uncovering the structural underpinnings of political discussion networks: Evidence from an exponential random graph model: Underpinnings of political discussion networks. *Journal of Communication*, *65*(1), 146–169. doi:10.1111/jcom.12140

Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, *60* (3), 556–576. doi:10.1111/jcom.2010.60.issue-3

Sunstein, C. R. (2006). *Infotopia: How many minds produce knowledge*. New York, NY: Oxford University Press.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. *The Social Psychology of Intergroup Relations*, *33*, 47.

Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review*, *109*(3), 451–471. doi:10.1037//0033-295X.109.3.451

Turner, J. C. (1975). Social comparison and social identity: Some prospects for intergroup behaviour. *European Journal of Social Psychology*, *5*(1), 1–34. doi:10.1002/ejsp.2420050102

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, *42*, 35–67. doi:10.2307/2393808

Varnelis, K. (2012). *Networked publics*. MA/MIT Press, Cambridge.

Watts, D. J., & Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of Consumer Research*, *34*(4), 441–458. doi:10.1086/518527

Weeks, B. E. (2015). Emotions, partisanship, and misperceptions: How anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation. *Journal of Communication*, *65*(4), 699–719. doi:10.1111/jcom.12164

Wilson, E. O. (2012). *The social conquest of earth*. W.W. Norton & Company. Liveright Pub. Corp, New York.

Xu, B., Huang, Y., Kwak, H., & Contractor, N. (2013). Structures of broken ties: Exploring unfollow behavior on Twitter. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (pp. 871–876). New York, NY: ACM. doi:10.1145/2441776.2441875