How Communication Can Make Voters Choose Less Well

Ulrike Hahn[1], Momme von Sydow[2] & Christoph Merdes[2]

[1]Department of Psychological Sciences, Birkbeck, University of London, Malet Street, London, WC1E 7HX, UK

[2]Munich Center for Mathematical Philosophy, Ludwigstr. 31, 80539 Munich, Germany

Preprint of

Article DOI: 10.1111/tops.12401

Journal: Topics in Cognitive Science (TopiCS)

25 December 2018 (online first)

2019 print

Corresponding author:

Ulrike Hahn, Department of Psychological Sciences, Birkbeck, University of London, Malet Street, London, WC1E 7HX, UK, u.hahn@bbk.ac.uk

word count: 4,500

Abstract

With the advent of social media, the last decade has seen profound changes to the way people receive information. This has fueled debate about the ways (if any) changes to the nature of our information networks might be affecting voters' beliefs about the world, voting results, and, ultimately, democracy. At the same time, much discussion in the public arena in recent years has concerned the notion that ill-informed voters have been voting against their own self-interest. The research reported here brings these two strands together: simulations involving agent-based models, interpreted through the formal framework of Condorcet's (1785) Jury Theorem, demonstrate how changes to information networks may make voter error more likely even though individual competence has largely remained unchanged.

Keywords: vote aggregation; Condorcet jury theorem; agent-based modelling; voting; communication

Introduction

Recent political developments in the Western world have prompted renewed focus on questions of voter competence. To some observers at least, significant numbers of voters seem to vote for candidates (and with them policies) that go against their own self-interest, prompting a revival of this longstanding issue in political science (Sears et al., 1980; Weatherford, 1980) within the contemporary popular press (see e.g., Zeitz, 2017; Brooks, 2017). With respect to voter competence, a case has been made that powerful long-term developments in the US, in particular, have led to a rejection of the very notion of expertise (Nichols, 2017). At the same time, however, political scientists are quick to point out that there is evidence to suggest that voters have never seemed particularly well-informed (for the debate about voter ignorance see e.g., Downs, 1957; Nannestad & Paldam, 1999; Sanders, 2000; Silva & da Silva Costa, 2006). Of course, voters may have interests and motivations, and hence reasons, that differ from those observers impute. In particular they may have reasons other than economic self-interest or other overtly self-related motivations (Cramer, 206; Tajfel & Turner, 1986; Turner, Oaks, Haslam & McGarty, 1994; Efferson, Lalive, & Fehr, 2008).

However, as will be explored in this paper, it is entirely possible that voters have always been comparatively poorly informed, but that there are nevertheless changes in the links between those individual information levels and aggregate outcomes. Specifically, the outcome of a group vote such as an election has properties that are distinct from those of the individual votes. And those properties are not just determined by the intrinsic characteristics of the individual votes but also the relationships between voters. If those change, so may the 'quality' of the voting outcome.

The present paper explores this through agent-based modelling set in the contexts of formal results on voting within the framework of Condorcet's (1785) jury theorem. It is demonstrated how changes to inter-agent communication (prompted, for example, by the advent of social media) may (negatively) effect the accuracy of collective judgement.

Background: Vote Aggregation

Much of the research on votes and the aggregation of votes has been focussed on *preferences*: in choosing between political candidates and party platforms, voters are expressing valuations. It is well-known since Arrow (1957) that the aggregation of preferences may lead to outcomes viewed as undesirable by most or even all of those expressing them.

Preferences are not our focus here. Rather our concern is with the antecedent facts on which those preferences are based, particularly in a time that has been called 'post-factual politics': are "conservatives the party of economic competence", is Hillary Clinton ”crooked", a "wall street shill" or a "war hawk", is Donald Trump a "successful businessman" or "suffers from dementia", and can "Brexit deliver an extra 350 million a week to Britain's national health service"?

These are not values, but propositions about the world; they are putative facts that may be true or false and our candidate preferences, in turn, depend on them. Where the perception of voters voting against their own self-interest is voiced, the assumption is typically that these voters are mistaken about the relevant antecedent facts.

Setting aside the empirical question of whether such mistakes have, in fact, played a causal role in recent electoral suprises, we explore here mechanisms by which such mistakes could plausibly occur with greater frequency, despite the fact that individual voters have not changed (or changed much) with respect to how well informed they are.

The key to this lies in the fact that voting, like other forms of judgment aggregation, may give rise to 'wisdom of the crowd' effects (see e.g., Surowiecki, 2004; Page, 2008). Whether it does or not, however, hinges crucially on the relationship between voters, and not just their individual competence. If this has fundamentally changed, then collective perceptions of putative 'facts' may become less accurate, even though voters themselves taken individually are no more poorly informed than before. The Condorcet jury framework provides a formal framework for understanding such effects.

Condorcet's (1785) Jury Theorem

Condorcet's (1785) jury theorem gives grounds for optimism concerning democracy by showing that majority voting may be a powerful tool for judgement aggregation. The theorem shows that given two mutually exclusive alternatives, such as the truth or falsity of a claim, a group verdict based on simple majority vote may outperform the individual judges in terms of accuracy and, as groups size increases, converge on the truth. Condorcet's basic result assumes n voters whose choices are independent of one another, and a probability p that each voter will pick the correct alternative which is assumed to be the same for all voters. If that probability is greater than .5 (assuming prior odds for the alternatives that are even), then the probability that the group choice, P_N, will be correct, will not only be higher than p (i.e. the group verdict will be more accurate than the individual voters), but it will increase rapidly with group size N, and will approach infallibility in the limit.

This power of majority voting obtains regardless of *how much* the voters know, as long as they know something (if their accuracy is at chance, i.e. $p = .5$, then the group verdict too will remain equal to tossing a coin; and, of course, if they are systematically biased against the right option, i.e., $p < .5$, then the reverse holds: P_N will be even lower).

Moreover, the rate of convergence to the asymptotic limit of "infallibility" is quite fast: for individual accuracy of $p = .08$, the probability that the majority vote of just 13 such voters is correct is greater than .99 (see Grofman, Owen & Field, 1983).

However, in the real world, the independence assumption will, more often than not, be unrealistic. People's judgments will be correlated because they share common information, because they communicate with each other, because they follow supposed 'experts', opinion leaders, or because they belong to certain schools of thought. In the limit, non-independence can mean that a group verdict is not really a group judgment at all, because it simply reflects a single opinion: Imagine, for example, a case where group members base their verdicts exclusively on the judgment of a single expert. In this case the groups' judgment will be no more (or less) accurate than that single expert.

Even if individuals base their judgments only *partly* on that expert, however, it will decrease the group's accuracy relative to what it would have been if votes were independent. To illustrate, assume, once again, equal levels of individual competence, p, for all group members (including the group's opinion leader) should they cast their votes independently, but assume also that group members have a certain probability d, of deferring to an opinion leader by simply adopting that leader's judgment. Then group competence will again collapse to that of the opinion leader as soon as d is sufficiently large relative to p. To provide an example by Grofman et al. (1983), if d = 0.2 and p = 0.6, then the expected value of P_N is 0.6, *regardless of group size N*. In other words, the group majority is only exactly as competent as the leader, since the leader's voting bloc will be likely to determine the outcome of the vote.

Of course, in the more realistic case of unequal competence, benefits to overall accuracy may ensue if the opinion leader is more accurate than some (or even all) of the other group members. However, there too the benefit in accuracy will be diminished by the costs of non-independence.

Ladha (1992), provides a more general, and more realistic version of Condorcet's theorem, that incorporates both differences in individual competence and non-independence of voters. Amazingly, whether or not the ampliative effect of group judgement holds or not, is determined only by the *average* probability of correct responding, $\bar{p}$ within the group, and by the *average* level of independence, $\bar{r}$. Specifically, Ladha shows that it is possible to calculate a threshold, $T(n, \bar{p})$, such that for a given group size n and mean competence, $\bar{p}$, levels of average interdependence below that threshold will guarantee that the overall group accuracy (P_N) will be higher than the average accuracy of the individual group members ($\bar{p}$):

if $\bar{p} > .5$, and $\bar{r} < T(n, \bar{p})$ then $P_N > \bar{p}$

where,

$$T(n, \bar{p}) = \bar{p} - \frac{n}{n-1}\frac{(\bar{p} - .25)(1-\overline{p})}{\bar{p}} \qquad (1)$$

$r_{ij,}$ here, is the probability that voters i and j will vote simultaneously for the correct alternative (r_{ij} = *P(Vote_i = correct,* $\wedge$ *Vote_j = correct*), which will be equal to the product of the individual's probability of responding correctly, $p_i p_j$ if the votes are independent, greater than this if they are positively correlated, and smaller if they are negatively correlated.

Beyond the threshold of Eq. (1), the group verdict P_N will approach the correct answer with certainty as group size approaches infinity and mean dependence $\bar{r}$ approaches $\bar{p}^2$. In other words, despite some degree of non-independence, group 'infallibility', is still possible, as long as the inter-dependence is not too high.

As a consequence, individuals may contribute to overall group accuracy both by raising mean accuracy and by lowering mean dependence. Furthermore, because it is precisely the mean of individual competence and the mean inter-voter independence that matter, and not a minimal level of competence or independence that must be met by each group member, individuals may *improve* group accuracy even if their competence is considerably *below* the mean competence in the group as long as they sufficiently lower the mean dependence.

Conversely, adding further expertise to the group will only be beneficial if it does not also increase too much the average interdependence. Additional information that improves accuracy will be most effective where it is not shared information between group members. If its effects are too homogenizing it may actually decrease accuracy. Adding in the influence of an expert's opinion may thus be detrimental even if that expert is far more competent than the other group members if the expert's influence is too strong.

These formal results have direct implications for communicating groups: communication may *increase* individual accuracy, but it will also *decrease* diversity. So, whether or not communication helps or hinders collective accuracy depends on which of these changes is greater.

That these formal results have bearing on real groups can be seen from experimental studies manipulating communication empirically and measuring increases or decreases in individual and collective accuracy (e.g. Joensson, Hahn & Olsson, 2015; for an example of increasing individual, but decreasing collective accuracy, see Lorentz et al. 2011). As a consequence, changing significantly the size and structure of our everyday communication networks, in particular through the rise of social media such as Facebook and Twitter, could have a profound effect on the collective 'wisdom' of our votes. Even though voters may have been no more well-informed in the past, an electorate basing its individual votes on largely private information about each voters' local expertise ("this candidate knows something about small business", "this candidate knows something about farming" etc.) may exhibit a far more accurate factual picture of candidates collectively than a contemporary society where shared information on Twitter and Facebook gives rise to significant amounts of shared 'information' across voters. The present paper seeks to demonstrate this through agent-based simulations of a simple Condorcet-like voting paradigm.

Models & Simulations

Our modelling framework is based on the work of Olsson (2011, 2013) and involves naïve Bayesian agents that receive both information from the world and communicate with one another according to the structure of their communication network. At issue is the truth or falsity of a single proposition. At each time point, there is a probability of an agent receiving information from the world that states that this proposition is true (or false) with a given accuracy, p. In other words, as in the Condorcet theorems outlined above, p, indicates the probability that this piece of information correctly indicates the truth. At each time step, there is also a chance that agents will communicate with other agents with whom they have connections (with connections bi-directional in these simulations). Agents communicate the putative truth (or falsity) of the proposition in question based on their own degree of belief: if that degree of belief exceeds a certain threshold they communicate that it is true, otherwise they stay silent (in our first simulations in Figure 1 to 3 that threshold is a subjective degree of belief $p(claim) = .8$, and symmetrically, $p(not\ claim) = 1 - .8 = .2$, where agents believe a claim to be false). Finally, the agents are naïve Bayesian agents only (and not optimal Bayesians) because, as in the real world, they do not have knowledge of the full network topology (and hence information paths), nor do they have full knowledge of the accuracy of their sources.

The agents in our model try to estimate the reliability of their sources on the basis of the extent to which the evidence they receive (from either world or others) matches their present beliefs. This strategy, which has been called 'expectation-based updating' (Collins et al, 2018) intuitively runs as follows: On receipt of evidence, they revise not just their belief in the hypothesis, but also in the reliability of the respective source based on the match between evidence and current belief (for formal details see Olsson, 2011). In effect, the simple logic of this kind of strategy runs like this: if you say to me something that I think is unlikely to be true, I will nevertheless increase my belief in what it is you are asserting, but I will also decrease my belief in your reliability. On hearing from someone that the Earth is flat, this strategy will make one think that this is a bit more likely to be true, but it will also make one think that person is less reliable than previously thought. This strategy not only seems intuitive, there is also experimental evidence for its use in even very simple contexts of testimony (Collins et al. 2018) and philosophers have considered it to be a rational, normative solution to the problem of determining the accuracy (reliability) of unknown sources (Olsson, 2011, 2013; Bovens & Hartmann, 2003). For robustness, and to isolate the impact of this strategy, we contrast these agents (who we refer to "update agents" in the remainder), with even simpler agents, who simply trust others in the sense that they assign a fixed degree of (moderate) accuracy to all sources (p = .66, "fixed trust agents"). For detailed examination of the utility, characteristics and problems of both of these strategies vis a vis a single source, see Hahn, Merdes & von Sydow (2018).

The agents in our simulation communicate within a small world network (see Watts & Strogatz, 1998) – a network structure found in many real-world social and biological networks that is characterized by short paths despite comparatively low link density. Crucially, to isolate the effects of communication within a social network, our simulations include, for each society of communicating agents, an equal-sized group of "shadow agents". Each of these shadow agents tracks one of the agents within the network of communicators by receiving all of the same information from the world as that agent, but without participating in any of the communication.

In the simulations reported here, each society is run for 50 times varying across simulations some of the key parameters in the model to explore the impact of communication. In keeping with the relationship between number of voters and accuracy of the majority vote set out in the Condorcet framework, we varied the network size in 21 steps from N = 4 to N = 100. Second, we also examined three levels of individual accuracy of the voters, by setting the probability that the information they receive from the world is correct to either p = .55, p = .66, or p = .75.

[1] The model formalizes the perceived accuracy of a source (trust) using a Beta distribution. We use a trust prior with $Mean(\theta)$ = .66, formalizing a plausible default assumption of moderate accuracy with confidence $Beta(2,1)$. For the fixed-trust agent perceived accuracy is fixed to .66. We assumed no initial preference for H or Non-H, i.e. $P(H)$ = .5, but in order to be able to detect potential effects of asymmetries of H and non-H, we set H to actually be true for 60% and false for 40% of the runs.

Results

Each data point in our results reflects the average across 100 runs of a society with the given parameter values.[1] Our results compile the following descriptive statistics: at the end of each run over the 50 time steps, individuals in a society assess their degree of belief in the proposition and cast a ‘vote’ on its truth or falsity: if their degree of belief is > .5 they vote “true”, if it is < .5 they vote “false” and if it is .5 exactly, their vote is split. To compute individual accuracy, *p* (which will deviate from the quality of the information received from the world, both because individuals may receive more than one piece of such evidence, and because networked agents also receive communication), we calculate the proportion of correct votes for that individual across simulation runs. At the same time, the individual votes are counted to establish the majority vote for the society as a whole. To determine group accuracy, we calculate the proportion of correct majority vote outcomes across the 100 simulation runs for a given set of parameters.

Figure 1 shows the key results assuming that everyone participates in voting. Dashed lines represent the individual accuracy and solid lines the accuracy of the group verdict.

The three different colours in each plot represent the different levels of quality for evidence received from the world, and the panels show results for the network of communicating agents (left panel) and their non-communicating shadow agents (right panel). As can be seen from the fact that the solid lines lie above those of

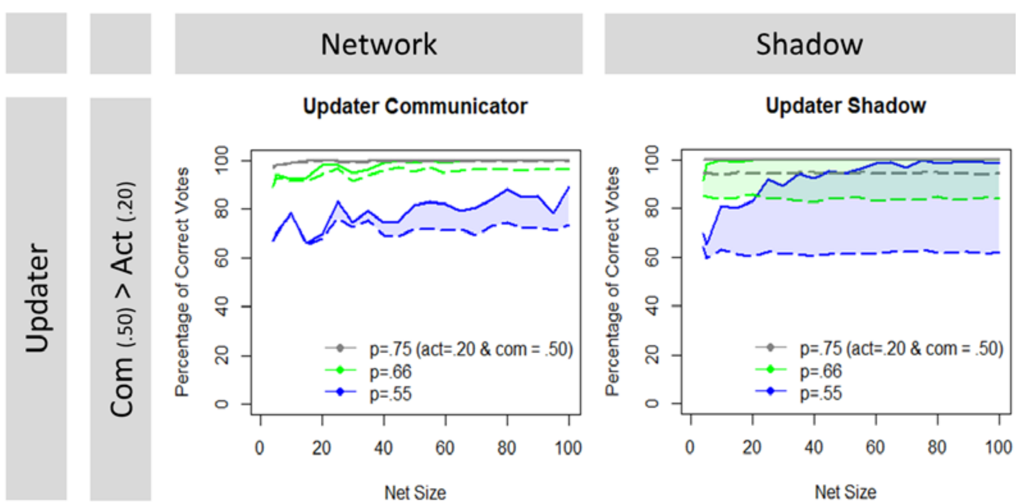

Figure 1: Displayed is the percentage of votes correct across 100 simulation runs. Dashed lines represent individual accuracy, solid lines represent collective accuracy (accuracy of the majority vote). The left panel shows the network results, the right panel the results for otherwise matched, but non-communicating agents. The different colours identify levels of accuracy for information coming from the world. Parameters “act” and “com” are explained in subsequent text below.

their respective dashed counterparts in all cases, the collective vote shows a clear benefit of aggregation in line with the Ladha (1992) results. However, the size of that accuracy boost for the collective is considerably reduced among the networked, communicating agents. Although the percentage of individually correct votes is slightly higher for the communicating agents, paradoxically the results for the average collective level are clearly lower. This deficit of the communicating agents is seen both in the slower rise in the accuracy of the collective vote as a function of group size, and in the levels of accuracy reached: where evidence quality was lowest (*p* = .55), accuracy was almost 20 percentage points lower even for the largest group. This decreased collective performance holds even though individual accuracy rises with communication, because diversity (and the information that diversity contains) is lost.

If communication reduces independence, then one should expect greater accuracy of collective voting where there is more evidence from the world, relative to evidence coming from communication.

To explore this, we also varied the degree of activity from the world ("*act*"), that is the probability of receiving data from the world in a given time step, relative to the probability of communication ("*com*"). Whereas the results shown in Figure 1 are based on values of *act* = .2 and *com* = .5, Figure 2 shows the corresponding results for probabilities *act* = .5 and *com* = .2. As expected, more evidence from the world leads not only to greater individual accuracy, but the greater relative proportion of 'evidence' coming from the world as opposed to communication, means that inter-dependencies are reduced, and a greater accuracy boost through voting is observed.

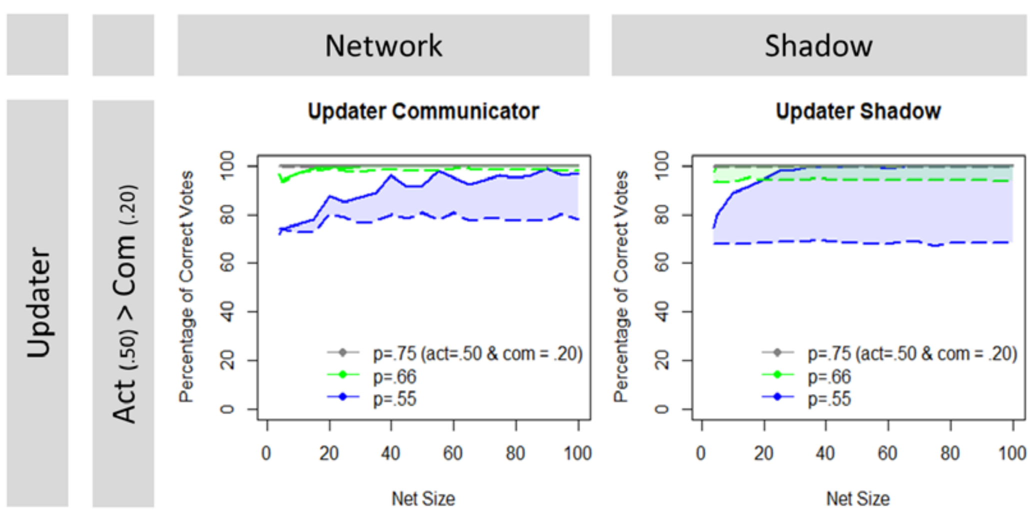

Figure 2: Corresponding results (as before, dashed line = individuals, solid line = collective) having swapped the probabilities of receiving information from world and via communication in the simulations of Fig. 1.

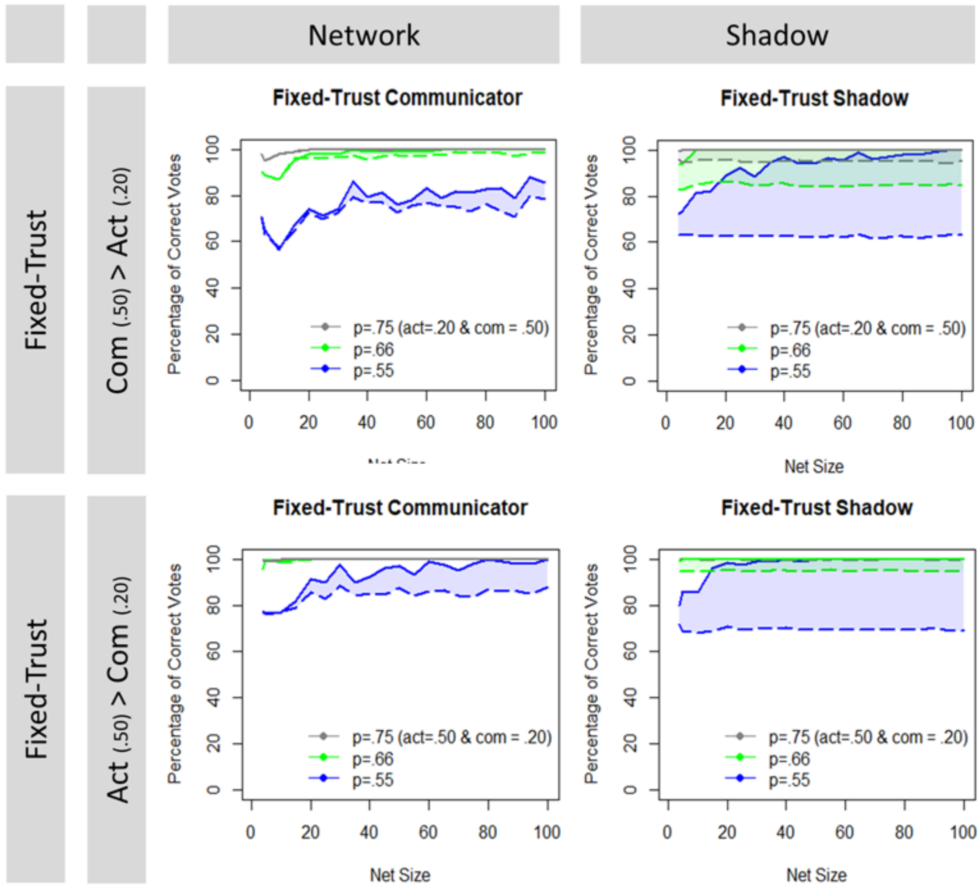

Figure 3: For the fixed-trust agent the top row shows the results for act = .2, com = .5 (cf. Fig. 1), the bottom row shows act = .5, com = .2 (cf. Fig 2). Dashed lines represent percentage correct for individual votes, solid lines for the majority vote.

These findings are mirrored in the results for "fixed trust" agents in Figure 3. There is virtually no difference between the "update" agents and the "fixed trust" agents. That fixed trust agents are no less accurate than the agents who try to estimate reliability seems counter-intuitive and surprising, but the present findings with networks of communicating agents match those for single agents in Hahn et al. (2018).

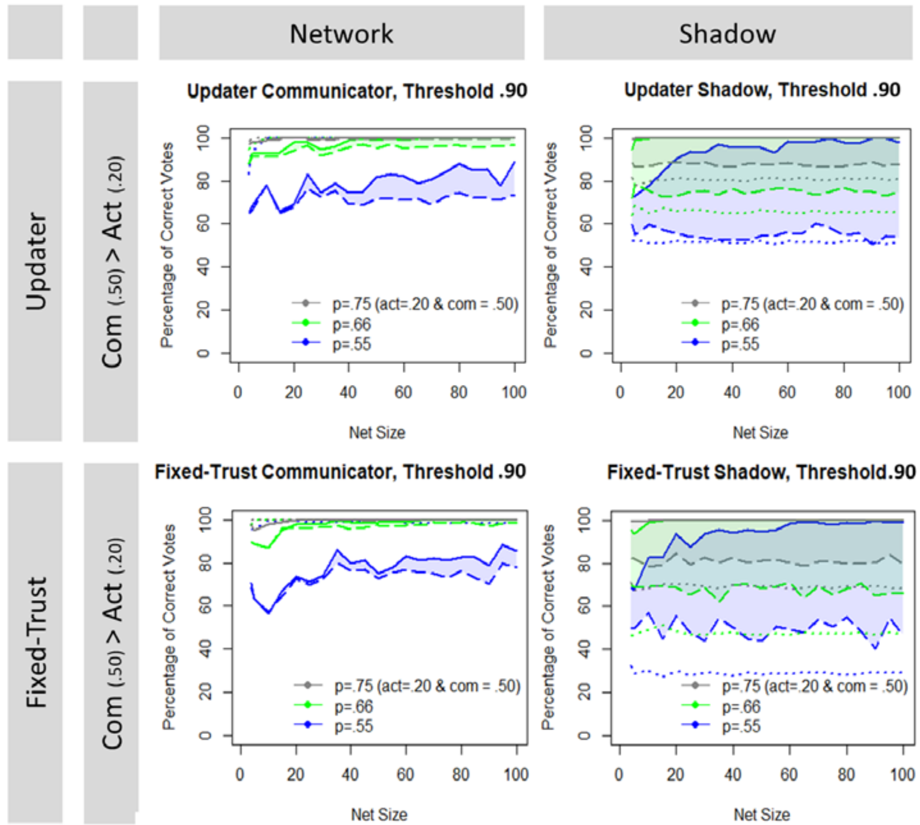

Figure 4. The figure shows the collective (solid line) and individual accuracies (dashed line) in the simulated populations of Fig. 2 and Fig. 3 top row, when only those agents vote whose voting threshold exceeds .9/.1. The additional dotted lines show % turnout of the vote.

Voting Thresholds

We also considered the implications of the fact that, in real world contexts, agents who are not sufficiently convinced either way may simply abstain from the vote. To examine this, we let only those agents vote, whose degree of belief exceeded a specified certainty threshold (.9/.1, respectively). The results are shown in Figure 4. The dotted lines represent the resulting turnout of the vote. As can be seen, the voting threshold has no impact on the *communicating* agents; their beliefs are already extreme enough that virtually all of them vote. It does, however, affect the non-communicating shadow agents and, here, differences between update and fixed-trust agents emerge: voting proportions are higher for updating agents at all evidence levels. What is common to both sets of shadow agents, however, is that there is almost no effect on collective accuracy, but a significant drop in mean *individual accuracy* if less extreme agents fail to vote (cf. Figure 1 and 2). This suggests those agents hold useful information, even though the information of the remaining agents is sufficient to boost collective accuracy. Further exploration of voting thresholds confirming these findings with other parameters, can be found in the online Supp. Mat.

Those simulations vary both the voting threshold and the threshold of assertion for communicating agents (that is, the degree of belief they need to surpass before they are confident enough to assert a claim and to communicate their belief to others). The results tentatively suggest that, in the context of our model at least, it has a much larger impact whether 'low confidence voters' participate in communication, than whether or not they ultimately vote. The generality of that relationship seems an important topic for future research.

Conclusions

In recent years, the advent of social media has seen a fundamental transformation in how we receive and communicate information: for example, 67% of US citizens now receive news via Facebook (Shierer & Gottfried, 2017), and the size and topology of people's effective communication networks has changed (on the topology of Facebook, see Ugander et al. 2011). Needless to say, most of the complexity of voter relevant communication is *not* captured in our simulations. However, the power of the Condorcet framework lies precisely in its abstraction: it is only the resultant individual accuracies and the degree of dependence that ultimately matter for collective accuracy; the processes by which these have been arrived at are ultimately irrelevant. What matters, then, is whether changes to our information networks have affected voter inter-dependence, as demonstrated in our simulations. It seems almost certain that they will have had *some* effect: so, yes, voters could collectively now be more ignorant than before even if individual accuracy has largely remained unchanged.

Acknowledgments

This paper is based on work presented at the 40th Cognitive Science Conference 2018 in Madison, WI, (US). The research project was funded by the Humboldt Foundation's "Anneliese Meier Research Award" to Ulrike Hahn.

References

Bovens, L., & Hartmann, S. (2003). *Bayesian Epistemology*. Oxford University Press.

Brooks, D. (2017) What's the matter with Republicans? https://www.nytimes.com/2017/07/04/opinion/republicans-government-programs.html?_r=0

Cramer, K. J. (2016). *The politics of resentment: Rural consciousness in Wisconsin and the rise of Scott Walker*. University of Chicago Press.

Collins, P.J., Hahn, U., von Gerber, Y., & Olsson, E.J. (2018). The Bi-directional Relationship Between Source Characteristics and Message Content. Frontiers in Psychology-Cognition, DOI: 10.3389/fpsyg.2018.00018

Downs, A. (1957) An economic theory of democracy. Harper & Row, New York.

Efferson, Ch., Lalive, R. & Fehr, E. (2008). The Coevolution of Cultural Groups and Ingroup Favoritism. *Science, 26*, 1844–1849.

Grofman, B., Owen, G., & Feld, S. L. (1983). Thirteen theorems in search of the truth. *Theory and Decision*, *15*(3), 261–278.

Hahn, U., & Harris, A. J. (2014). What does it mean to be biased: Motivated reasoning and rationality. *The psychology of learning and motivation*, *61*, 41–102.

Hahn, U., Merdes, C. & von Sydow, M. (2018). How Good is Your Evidence and How Would You Know? *Topics in Cognitive Science*.

Jonsson, M. L., Hahn, U., & Olsson, E. J. (2015). The kind of group you want to belong to: Effects of group structure on group accuracy. *Cognition*, *142*(C), 191–204.

Ladha, K. K. (1992). The Condorcet Jury Theorem, Free Speech, and Correlated Votes. *American Journal of Political Science,* 36*(3), 617–634.*

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. Proceedings of the National Academy of Sciences, 108(22), 9020–9025.

Nannestad, P., Paldam, M., 1999. What do voters know about the economy? A study of Danish data, 1990–1993. Electoral Studies 19(2).

Olsson, E. J. (2011). A simulation approach to veritistic social epistemology. *Episteme*, *8*(02), 127–143.

Olsson, E. J. (2013). A Bayesian simulation model of group deliberation and polarization. In *Bayesian argumentation* (pp. 113-133). Springer Netherlands.

Page, S. (2008). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies (New Edition)*. Princeton University Press.

Sanders, D. (2000). The real economy and the perceived economy in popularity functions: how much do voters need to know?: A study of British data, 1974–97. *Electoral Studies*, *19*(2), 275-294.

Sears, D., Lau, R., Tyler, T., & Allen, H. (1980). Self-Interest vs. Symbolic Politics in Policy Attitudes and Presidential Voting. *American Political Science Review*, *74*(3), 670-684.

Shierer, E. & Gottfried, G. (2017). News Use Across Social Media Platforms 2017. http://www.journalism.org/2017 /09/07/news-use-across-social-media-platforms-2017/

Silva, E. G., & da Silva Costa, J. (2006). Are voters rationally ignorant? An empirical study of Portuguese local elections. *Portuguese Economic Journal*, *5*(1), 31-44.

Surowiecki, J. (2004). The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business. *Economies, Societies and Nations*, *296*.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behaviour. In S. Worchel & W. G. Austin (Eds.), *Psychology of Intergroup Relations* (pp. 7–24). Chicago, IL: Nelson-Hall.

Turner, J. C., Oaks, P. J., Haslam, A, McGarty, C. (1994) Self and Collective: Cognition and Social Context. *Personality and Social Psychology Bulletin*, *20*(5), 454–463.

Ugander, J., Karrer, B., Backstrom, L., & Marlow, C. (2011). The anatomy of the facebook social graph. *arXiv preprint arXiv:1111.4503*.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440-442.

Weatherford, S. M. (1983). Economic Voting and the "SymbolicPoltics" Argument: Reinterpretation & Synthesis. *American Political Science Review*, *77*(1), 158-174.

Wineburg, S., & McGrew, S. (2016). Why Students Can't Google Their Way to the Truth. *Education Week*.

Zeitz, J. (2017). Does the White Working Class Really Vote Against Its Own Interests? https://www.politico.com/magazine/story/2017/12/31/trump-white-working-class-history-2162.

Supplemental Materials (online)

To Hahn, U., von Sydow, M., & Merdes, C. (2019). How Communication Can Make Voters Choose Less Well. *Topics in Cognitive Science.*

The following figures summarize the results of *additional simulations* in order to provide the reader with some sense of the generality of the results presented in the main text. While they in no way constitute an exhaustive search of the parameter space of the model, they do show that the results reported are not limited to one narrow set of initial parameters. Specifically, we replicated the simulations reported in the main text with an even more extreme balance between evidence from the world and communication as these would be expected to draw out even more clearly the underlying points. To this end, we simulated populations where the probability of evidence from the world on a given time step was *p(act)* = .1 and the probability of communication (once the agent's threshold of assertion —TOA— was passed) was *p(com)* = .9, whereas the figures reported in the main text either used *p(act*) = .2 and *p(com)* = .5, or *p(act)* = .5 and *p(com)* = .2.

We also varied the TOA: whereas the simulations in the main text were based on a TOA of .8, we examined the impact of both lower, TOA = .6, and higher, TOA = .9, values below. We combined this with three possible levels of the voting threshold, TOV: no threshold = everybody votes; TOV = .9/.1, that is, only voters with beliefs more extreme than .9 and those more extreme than .1 voted (as in the simulations of the main text), and a more relaxed threshold of TOA = .6/.4.

All other parameters were kept to the values reported in the main text (e.g., run length = 50 time steps), and, as in the main simulations, each data point represents an average of 100 runs.

The *results* of these simulations are displayed in Figure Panels S1 to S3. These additional simulations confirm the key patterns reported in the main text: Communication damages the "wisdom of crowds", significantly reducing the accuracy gain of the collective vote (apparent in all three figures S1, S2, and S3). This holds for all twelve pairs of graphs comparing network and shadow agents presented in the three panels. Regardless of whether the average individual accuracy is higher for the communicating agents or for the shadow agents, collective accuracy is always higher for the shadow agents, as is the accuracy boost from averaging (wisdom of crowds effect).

Once again, the introduction of voting thresholds (Figure S1 and S2) materially affects only the voting turnout of the non-communicating agents (the shadow agents); in other words, the communicating agents already have sufficiently extreme beliefs to ensure voting. Across the board – and perhaps surprisingly – the voting threshold clearly lowers the average *individual* accuracy for the shadow agents, but again has little impact on *collective* accuracy (compare the top four and the bottom four panels in S1, and compare the top four and the bottom four panels in S2).

The TOA, however, has a significant impact on accuracy for the communicating agents but not for the shadow agents, both individually and collectively: For the communicating agents, lowering the TOA (Fig. S2 relative to S1) not only lowers both individual and collective accuracies, but there is also no longer any increase in individual accuracy as a result of communication (contrast the top four panels in S1 with the top four of S2 for the case of everyone voting, and see S3 for the same comparisons with a TOV of .6). Importantly, for the collective vote, altering who votes has much less effect than altering who speaks in prior communication

Finally, the main difference between the update and the fixed-trust strategy lies in the proportion of the respective populations whose belief exceeds a TOV of .9 (mainly in the shadow agent populations, see right-hand panels S1 and S2). No such difference is apparent when the TOV is dropped to .6, indicating that the updating agents form more extreme beliefs. This is in keeping with the fact that the update strategy has ampliative effects, whereby the belief-congruence of evidence modulates the reliability of the source (see Hahn et al., 2018).

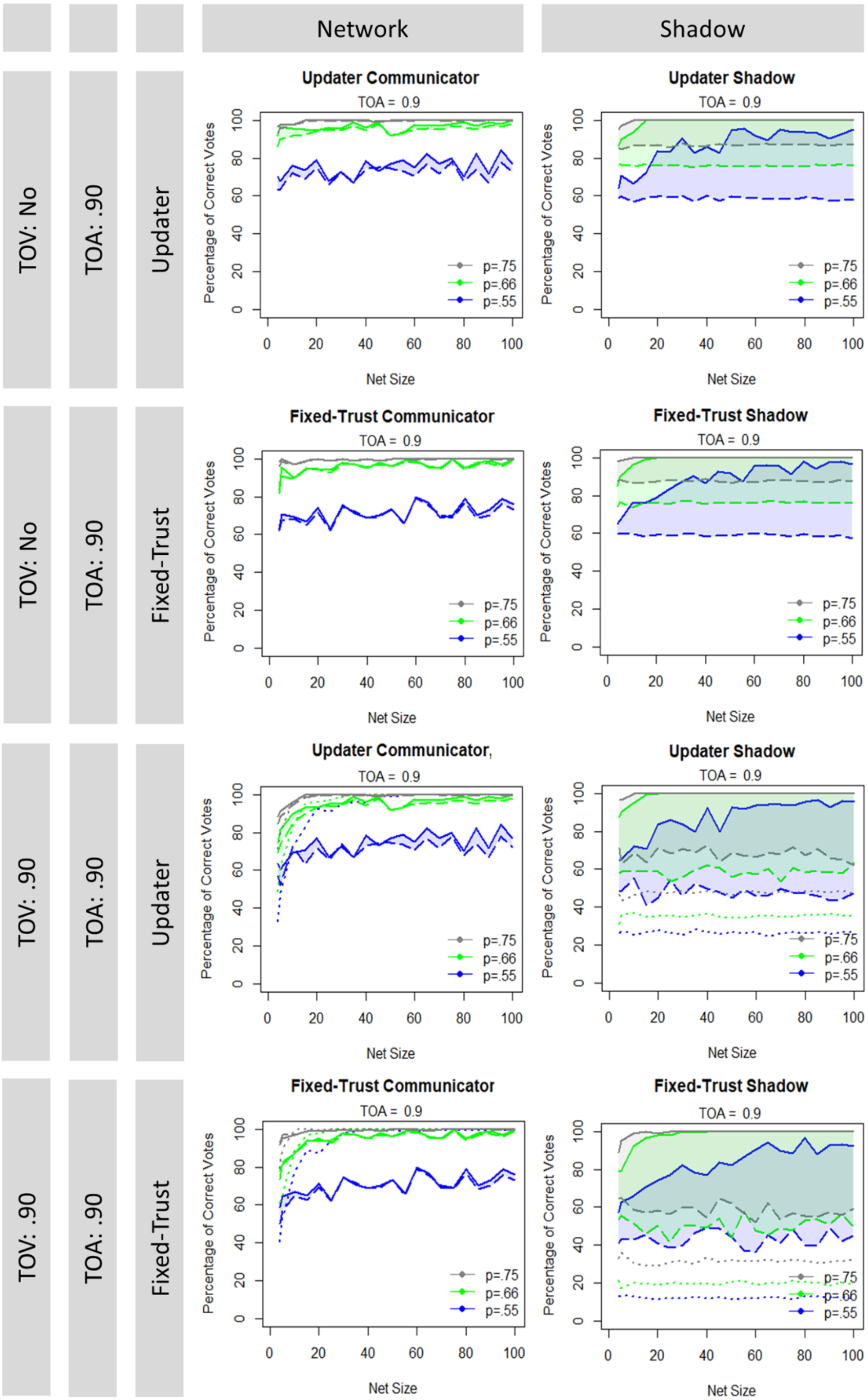

Figure S1. The results for communicating (network) and non-communicating (shadow) agents (with act = 10 and com = 90), contrasting the effects of no-voting threshold (top four panels) with those of a Threshold of Voting (TOV) of .9 (bottom four panels), for a high Threshold of Assertation (TOA = .9). Dashed line = individuals, solid line = collective, dotted line = turn-out of the vote.

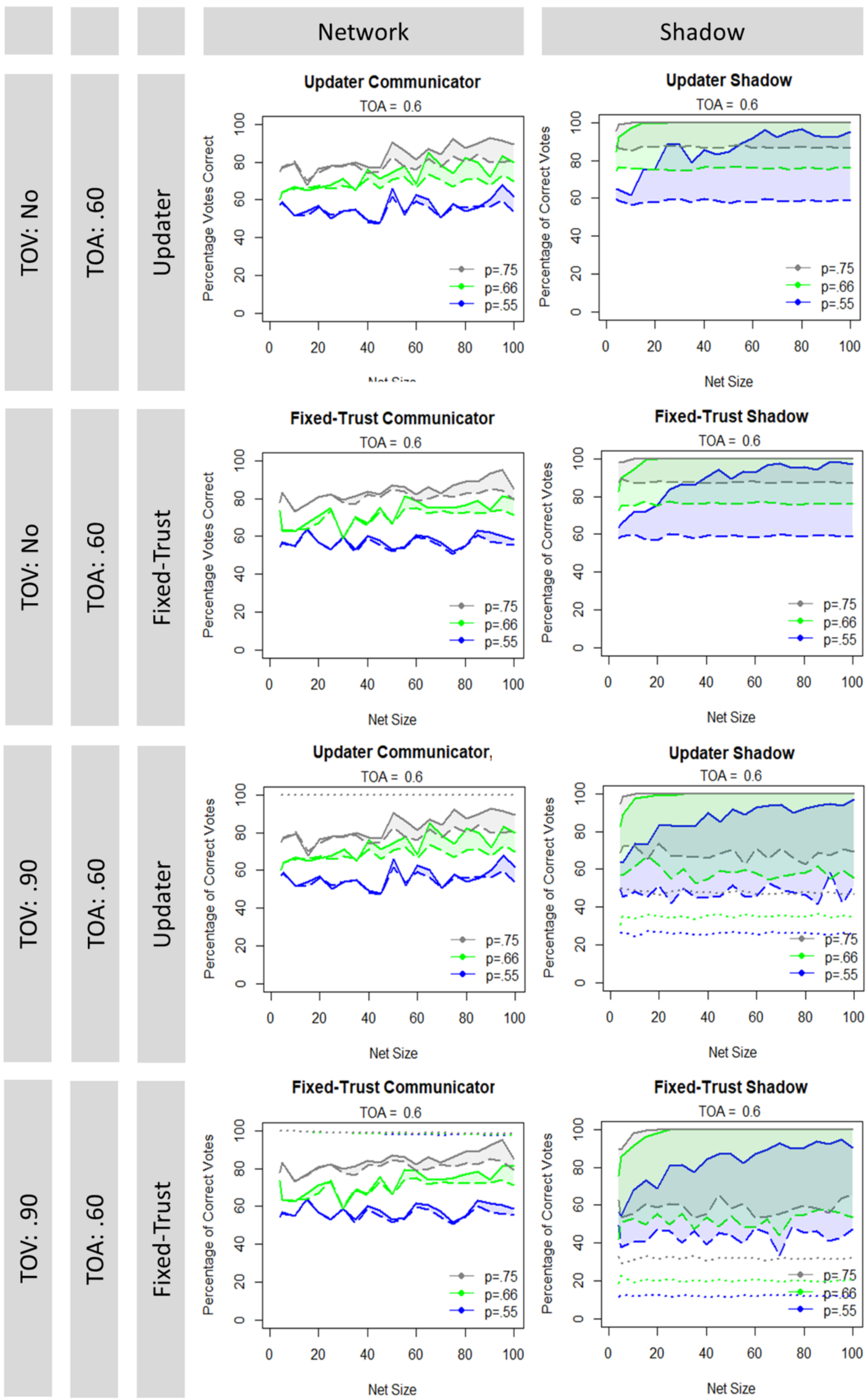

Figure S2: The results for network (left panels) and shadow agents (right panels) (with act = 10 and com = 90), contrasting the effects of no voting threshold (top four panels) with those of a TOV of .9 (bottom four panels), for a low TOA (.6).

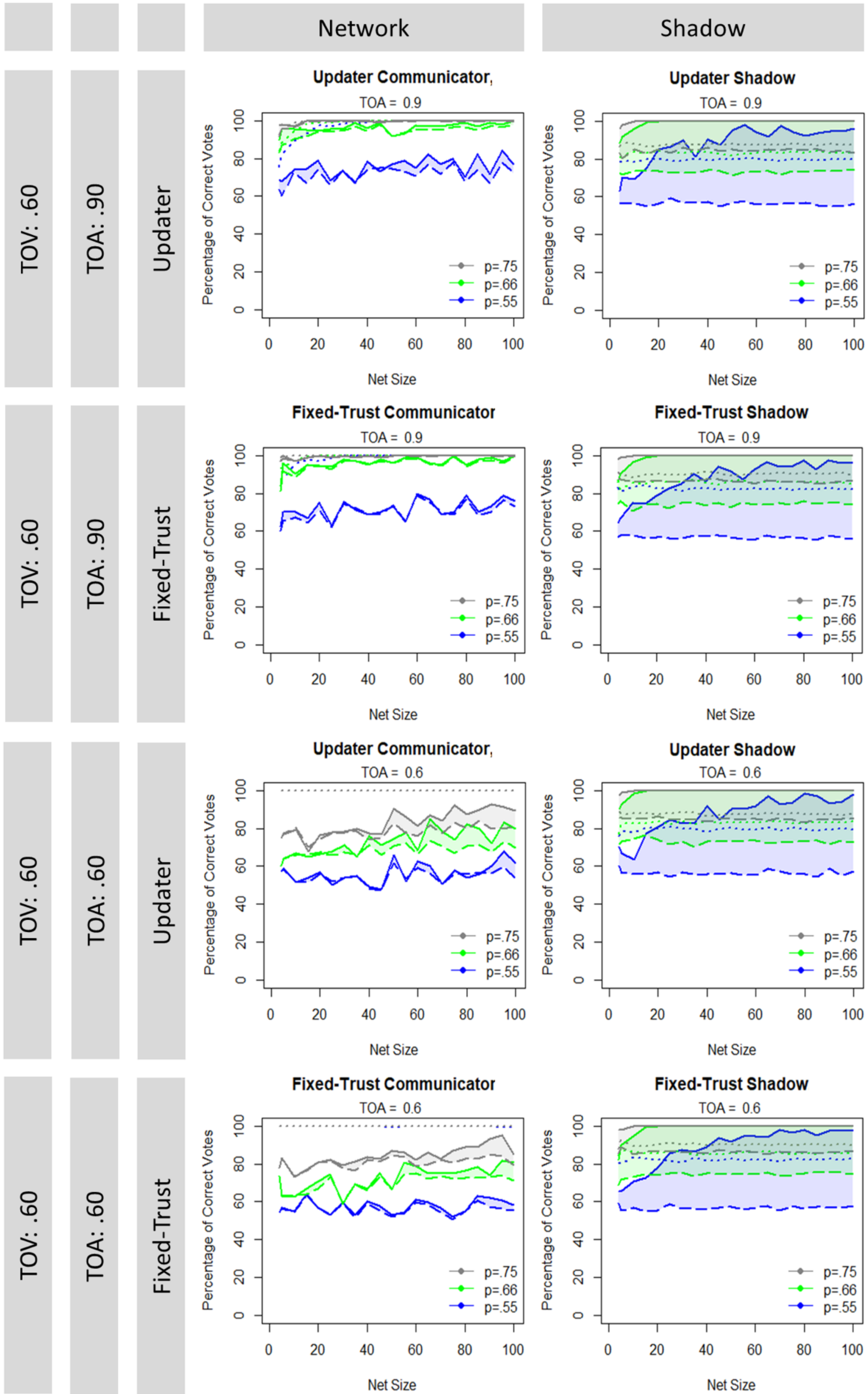

Figure S3: The results for network (left panels) and shadow agents (right panels) (with act = 10 and com = 90), contrasting differences in the TOA (.6 vs. .9, top two left-hand panels vs. bottom two left-hand panels respectively) in the context of a low voting threshold, TOV (.6). There is a clear effect of TOA. When the TOV is as low as .6, it has virtually no effect.