Discussion Paper Series

Institute-Wide Joint Research Project
Methodology of Social Sciences: Measuring Phenomena and Values

Are good leaders truly good?

Susumu Cato
University of Tokyo

Akira Inoue
University of Tokyo

May 9, 2022

E-22-001

Institute of Social Science, University of Tokyo
[https://web.iss.u-tokyo.ac.jp/methodology/en/dp/dp/]

Are good leaders truly good?

Susumu Cato
Institute of Social Science
University of Tokyo
7-3-1, Hongo, Bunkyo
Tokyo 113-0033
Japan
susumu.cato@gmail.com

Akira Inoue
Department of Advanced Social and International Studies / Graduate School of Arts and Sciences
University of Tokyo
3-8-1, Komaba, Meguro
Tokyo 153-8902
Japan
inoueakichan@g.ecc.u-tokyo.ac.jp

Abstract. This paper offers a new insight on the Condorcet Jury Theorem (CJT) in the theory of epistemic democracy. This theorem states that democratic decision-making leads us to correct outcomes under certain assumptions. One key assumption is the 'independence condition', which requires that voters form their beliefs independently when they vote. This paper examines the role of an opinion leader as an informational source, which potentially violates independence. We demonstrate that voters' beliefs may be correlated in the presence of the leader, and that the CJT can fail if the leader's opinions are reliable. This leads us to the following paradoxical observation: for epistemic democracy, good leaders may be bad, while bad leaders may be good.

Keywords: Condorcet Jury Theorem, independence, opinion leaders, public signals

Acknowledgments: We thank Lajos Brons for his helpful comments. This paper was financially supported by JSPS KAKENHI (20H01446 and 22H00598) and the Mitsubishi Foundation (ID201920011). This paper is prepared with support from the Institute of Social Science, University of Tokyo, and its institute-wide joint research project, 'Methodology of Social Sciences: How to Measure Phenomena and Values'.

1 Introduction

What is an essential characteristic of a good leader? One simple answer is that a good leader is an influential person whose opinions are likely to be correct. If this rough definition of good leaders—including opinion leaders—is accepted, then do you want good leaders in contemporary society? Many people might give an affirmative answer to this question. However, this study offers a counter-intuitive answer that sounds much like the witches in *Macbeth*: a good leader may be bad, and a bad leader may be good.

This study examines the role of opinion leaders in the theory of epistemic democracy, which has been recently developed by Goodin and Spiekermann (2018: Ch. 11). Specifically, we extend the canonical model for the Condorcet Jury Theorem (CJT) by assuming that opinion leaders provide 'public signals' to all voters. The CJT was first suggested by Black (1958) and received its standard formulation by Grofman (1975). In brief, this widespread theorem states that the larger the number of voters, the better the majoritarian outcome, and that the probability that the majoritarian outcome is correct approaches one, provided that the voters are competent.[1]

This theorem depends on the *independence condition*, which requires that the beliefs of voters are independent from each other, such that each voter contributes information without causally interacting with any other voter. This condition has met significant criticism. Most important in the present context, it has been pointed out that influential opinion leaders can push the electorate in a specific direction, thereby causing a violation of the independence condition and the CJT (Rawls 1971: 538; Grofman and Feld 1988; cf. Estlund 1994). Considering that something like this probably occurs in any society that has some kind of democratic decision-making, this is a serious concern.

Our argument in this paper offers a new insight on the influence of opinion leaders, focusing on their 'quality'. As alluded above, we say that the quality of an opinion leader is high (resp. low) if the probability that the leader's opinion is correct is high (resp. low); and the leader is considered good (resp. bad) if their quality is high (resp. low). We compare good opinion leaders with bad opinion leaders and investigate which of these are more likely to produce a good outcome if voters have equal access to their opinions. If voters know the quality of their opinion leaders, the degree to which they trust their leaders should be endogenously determined by voters based on that knowledge. This inference by voters leads to the apparently paradoxical

[1] A voter is competent in the sense required if the probability that they vote for the correct option is greater than fifty percent.

observation mentioned above. That is, the quality of leaders tends to be negatively correlated with the probability of success under majoritarian decision-making; this point is demonstrated in Bayesian and non-Bayesian models. Furthermore, by using a numerical simulation, we show that in the comparison of two good leaders, the better of the two would produce better outcomes because the CJT fails in either case.

This paper is organized as follows: Section 2 offers a benchmark model in which voters use Bayesian inference in the presence of a single opinion leader. In this model, the leader's opinion is understood as a public signal, which is acknowledged by all voters. Each voter is assumed to make a decision by comparing their private opinion with the leader's opinion. Section 3 extends this by relaxing the assumption of Bayesian reasoning by voters, resulting in a non-Bayesian CJT model that is compatible with the standard scenario used by Goodin and Spiekermann (2018) and others.

2 Baseline analysis

This section offers our basic observation on the role of an opinion leader. Specifically, we introduce a model with Bayesian inference in which each voter updates their belief as they gain relevant information and then make a choice to maximize their expected payoff based on the updated belief. It is important to notice that we do not claim the relevance of Bayesian inference in voting behavior; that is, we do think that it works as a benchmark that helps to understand belief formation in collective decision-making given opinion leader quality.

Let us first review how Bayesian inference works in the CJT framework. Consider a society with n voters who face a collective choice of policy-making. There are two social policy-related options: 'left' or 'right'. There are two and only two states of the world, 'L' ('left' state) and 'R' ('right' state), that are equally probable (i.e., both have probability 0.5). Every voter votes for either of the two policies. Hence, the voters must together choose right or left by collective decision-making. We assume that if left (resp. right) is chosen when the true state is L (resp. R), each voter gets one as their payoff. On the other hand, there is no payoff (i.e., zero) if they make an incorrect choice, that is, right (resp. left) is chosen when the true state is L (resp. R).[2]

Suppose that voters cannot know the true state directly. Instead, they can receive private signals, 'A' or 'B', through access to evidence or cue-taking. These private signals are private in the sense that voters cannot know other people's private signals. Furthermore, any voter's private

[2]The payoff structure can be generalized, but for simplicity we focus on the symmetric case.

signals are independent from the private signals of others. The probabilities of the two kinds of private signals depend on the true state: if the true state is L, then the probability of signal A is p (i.e., $P(A \mid L) > 0.5$) for each voter, while the probability of B is $1 - p$ (i.e., $P(B \mid L) < 0.5$). Similarly, if the true state is R, then the probability of signal B is p, while that of A is $1-p$ (i.e., $P(B \mid R) > 0.5$ and $P(A \mid R) < 0.5$). That is, signal A tends to be obtained by voters when the true state is L, while signal B is more likely when the true state is R. Notice that because the signals are private, as long as $p < 1$, different voters can get different signals.

For a voter following Bayesian reasoning, the question is what the conditional probability of L (or R) is, given their private signal. The conditional probability is calculated by utilizing *Bayes' theorem*:

$$P(\text{the true state is `}x\text{'} \mid i\text{'s signal is `}y\text{'}) = \frac{P(y \mid x)P(x)}{P(y)}.$$

If we apply this to the foregoing, we get the following equation:

$$P(\text{the true state is `}x\text{'} \mid i\text{'s signal is `}y\text{'}) = \frac{P(y \mid x)}{P(y \mid L) + P(y \mid R)}.$$

From this, we conclude that if a voter gets A (resp. B) as a private signal, they will believe that the true state is L (resp. R) with probability p and that it is R (resp. L) with probability $1 - p$. In order to maximize their expected payoff, voters with signal A (resp. B) vote for L (resp. R) because $p > 0.5$. Then, from the ex-ante viewpoint, each voter will choose the true state with probability p.[3] In this scenario, results of voting accord with the CJT: the larger the number of voters in society, the larger the probability that they will collectively get the true state. Moreover, the probability of getting the truth converges on one as the number of voters approaches infinity.

Next, we extend the previous scenario to include a *public signal* in addition to private signals.[4] A public signal is observed by *all* voters. In actual practice, this typically takes the form of an opinion leader reaching all voters by means of some influential media. Here we will assume that there is only one such public signal that all voters receive in addition to their private signal. This assumption implies that there is only one leader (or multiple leaders who completely agree in their opinions), making only one suggestion in a publicly influential manner.[5]

[3]This essentially implies that the assumption of competency follows as a consequence of Bayesian inference.

[4]See Austen-Smith and Banks (1996) for a pioneering analysis of a public signal. They do not assume the *sincerity condition*, one of the basic assumptions necessary for the CJT, which requires that each voter votes sincerely for the alternative that he or she believes is the right alternative. We assume sincerity. Austen-Smith and Banks (1996) examine strategic behaviors, but these are not our focus here.

[5]Nevertheless, it is not difficult to extend this to a case with multiple public signals.

Like the private signals (in both the previous case and this one), the public signal is either A or B. If the true state is L, the public signal is A with probability q (i.e., $P(A \mid L) > 0.5$), while it is B with probability $1-q$ (i.e., $P(B \mid L) < 0.5$). Similarly, if the true state is R, the public signal is B with probability q, while it is A with probability $1-q$ (i.e., $P(B \mid R) > 0.5$ and $P(A \mid R) < 0.5$). The key assumption here is that all voters receive the same signal. If q is high (i.e., the probability that the leader's opinion is correct is high), they are called a *good opinion leader*. If q is low (i.e., the probability they are called a *bad opinion leader*. The parameter q basically corresponds to the quality of a leader.[6]

We now consider the conditional probability of L (or R), given private signal y and public signal z. Using *Bayes' theorem* we get the following formula:

$$P(\text{the true state is `}x\text{'} \mid i\text{'s private signal is `}y\text{' \& the public one is `}z\text{'}) = \frac{P(y,z|x)}{P(y,z|L)+P(y,z|R)}.$$

From this, it can be easily inferred that if a voter's private signal coincides with the public signal, there is a high probability that these signals reflect the true state, and therefore, the voter votes rationally. However, a problem occurs in cases where there is no agreement between the two signals. Suppose that some voter gets A as their private signal and B as the public signal. If the true state is L, then the probability of the combination of these signals is $p(1-q)$, while it is $q(1-p)$ if the true state is R. In this case, which signal does the voter consider more likely to be true, (private) A or (public) B? Bayesian inference gives the following answer:

$$P(L \mid y = A \;\&\; z = B) = \frac{p(1-q)}{p(1-q)+q(1-p)},$$

and:

$$P(R \mid y = A \;\&\; z = B) = \frac{q(1-p)}{p(1-q)+q(1-p)}.$$

According to these formulas, L is more likely than R if p is larger than q, whereas R is more likely than L if q is larger than p. This seems reasonable because p and q represent the accuracy of private and public signals respectively. The voter's rational strategy is to follow the more accurate signal. This strategy is summarized in Tables 1 and 2.

Let us consider these votes and their implications. If the private signal is more reliable than the public signal, then each voter votes in accordance with the private signal. In this case, the

[6]In this baseline model, we distinguish two kinds of leaders (and their signals): if $q > p$ holds, then a leader is called a *good opinion leader*, on grounds that they know better than the citizens. If $q < p$ holds, they are called a *bad opinion leader*. Notice that the goodness (and badness) of leaders at issue is relative to the competence of each voter, and thus that, the ratio q/p represents the quality of a leader. However, this definition of the quality of a leader does not work in our extended model in Section 3.

	public A	public B
private A	vote for L	vote for L
private B	vote for R	vote for R

Table 1: the case where p is larger than q (bad opinion leader)

	public A	public B
private A	vote for L	vote for R
private B	vote for L	vote for R

Table 2: the case where q is larger than p (good opinion leader)

CJT mechanism functions: the greater the number of voters, the higher the probability of getting the true outcome. On the other hand, the CJT fails if the private signal is less reliable than the public signal. In this case, each voter follows the public signal, and there is, thus, a perfect correlation among people's votes, which violates the independence condition. In this case, we *derive* a correlation of votes rather than assuming it. Since the probability of public signal being right is q, the outcome of voting concurs with the true state with probability q, no matter how many voters there are. (And conversely, the outcome is wrong with probability $1 - q$.)

In summary of the foregoing, we find that, on the one hand, when voters have a bad opinion leader who is less reliable than their own opinion (provided that they know this to be the case), then a simple majority vote leads to the correct outcome if there is a sufficient number of competent voters, simply because the leader's opinion is ignored. On the other hand, when voters have a good leader (again, provided that they know this to be the case), the CJT fails because they follow the leader's opinion (violating the independence condition).

3 Robustness of our observation

The previous section shows that if voters are rational in a Bayesian sense, they will follow a good opinion leader, while they will vote in accordance with their own judgment when presented with a bad opinion leader. This condition of rationality might be considered a 'big if', because ordinary people are not (such) rational agents, and for that reason, the assumption of Bayesian rationality could be considered too strong for these findings to carry normative implications for collective decision-making in the real world. Indeed, our observation about the correlation between people's votes appears extreme. In the last scenario in the previous section, there is a threshold q^* beyond which the quality of the opinion leader is such that people's votes are perfectly correlated, while their votes are independent if q is smaller than q^* (i.e., leader quality is below the threshold). This sharp cut-off point and its extreme consequence do not correspond with the voting behavior of real people.

However, the principal aim of our argument is to demonstrate how the violation of the independence condition depends on the quality of opinion leaders. Even if the assumption of Bayesian reasoning is dropped, our main argument remains valid. In other words, our claim that good (resp. bad) leaders may produce worse (resp. better) outcomes is robust in the standard, non-Bayesian model of the CJT.

Here, we assume that people tend to pay heed to what good leaders say and not to what bad leaders say. We believe that this assumption is acceptable even for those who reject the assumption of Bayesian inference by voters. For example, if the quality q of the leader is 0.7, then each voter follows the leader's opinion with probability 0.1, while if q is 0.6, then each voter follows the leader's opinion with probability 0.05. This is not extreme and, indeed, is not far from reality.

Now, let us consider a more general model in the fashion of Goodin and Spiekermann (2018). Each voter independently makes a correct judgment with probability p if they are not affected by the opinion leader, but they respond to the signal from the opinion leader at r. Each voter then votes in accordance with the public signal with probability r, while they vote according to their own opinion with probability $1 - r$. Assume that r is an increasing function of the quality of the leader q, which tracks the probability of the opinion leader's judgments being correct. This implies that the better the leader, the more likely it is that voters follow them. The scenario with Bayesian reasoning is an extreme case hereof: if the leader is more reliable than private signals then r is 1, otherwise it is 0. In other words, if the assumption of Bayesian reasoning is relaxed and r is an increasing function of q with values between 0 and 1, then the above argument is consistent with non-Bayesian CJT models such as that by Goodin and Speakerman (2018: 165–9) where r is constant.

To see the implications of this new (non-Bayesian) scenario, consider the following three pairwise comparisons of leaders with different qualities.[7] First, in a comparison between a good and

[7]Mathematica, Version 12.0 (Wolfram Research, Inc. 2019) is employed to conduct simulation results shown in Figures 1–3.

a bad leader, the better the leader, the more likely the CJT fails, and conversely, the worse the leader, the more likely the CJT functions because voters tend to think little of the leader's opinion. This case is illustrated in Figure 1. Here, p is set at 0.55.[8] For a good leader, q is 0.8 and each voter follows the leader with probability 0.20. On the other hand, a bad leader succeeds with probability 0.6, but voters follow this leader with probability 0.05. One can see that the CJT fails when the leader is good, while it holds when the leader is bad. Hence, this is the same conclusion as in our observation in the previous section.

Second, we can compare two bad leaders. Because they are bad (and voters know this), the probability of people following their opinions is low, and the CJT holds for both leaders. We may say that the worse a leader among bad leaders, the more likely the result is 'approximately' better. In case of the less reliable (i.e., worse) leader the probability of voting correctly asymptotically approaches 1 more rapidly as the number of voters approaches infinity, which is shown in Figure 2. (The reason for this is that the worse a leader, the less likely it is that voters follow that leader, and thus the more likely it is that they follow their own, better opinions.) However, when the number of voters is small, a less bad leader is preferable, as shown in the part of the graph in Figure 2 corresponding to n less than about 70. Note that the number of voters is 139 when $n = 70$.

Third, we can compare two good leaders. As the CJT does not work for good leaders, this implies that it does not work for either of them, as illustrated in Figure 3. In this situation, the competency of leaders is crucial if n is (very) large. In that case, the better a leader among good leaders, the more likely the outcome is rational.

Interestingly, whether a good leader (i.e., a leader whose choice is more likely to be correct) actually produces optimal outcomes depends on the number of voters, and under certain conditions, a (relatively) good leader may produce worse outcomes. This is a novel discovery that was not observed in the Bayesian model in the previous section. As the figures show, our claim that good (resp. bad) leaders may produce outcomes that are more likely to be incorrect (resp. correct) is not negligible in the non-Bayesian CJT setting.

[8] We use the same competency level as that in Goodin and Spiekermann (2018: Ch. 11). As they point out, the threshold is $r^* = 0.09$. If $r < 0.09$, the CJT holds; otherwise, it fails.

4 Conclusion

This paper shows that voters' beliefs may be correlated in the presence of opinion leaders, and that the CJT can fail if the leader's opinions are (known to be) reliable. We thereby demonstrate that a paradoxical result holds in most cases: for epistemic democracy, a good leader may produce 'bad' outcomes (i.e., incorrect outcomes), and a bad leader may produce 'good' outcomes (i.e., correct outcomes).

It should be noticed that this argument also shows that the rationality of collective decisions is not only affected by the quality of opinion leaders (see Section 2), but also by the number of voters (see Section 3). In his inspiring work, Estlund (in Estlund et al. 1989: 1318–19) notes that this complex influence of opinion leaders has not been thoroughly studied. We believe that the results presented in this paper will contribute to the literature by exploring a difference between the Bayesian model, which provides a simple, base-line result, and a non-Bayesian model, which offers a certain complexity, and may thereby help reveal the sources of the complexity of leader influence observed by Estlund.

References

[1] Austen-Smith, D., and Banks, J.S. 1996. Information, aggregation, rationality, and the Condorcet jury theorem. American Political Science Review 90(1): 34–45.

[2] Black, D. 1958. The theory of committees and elections. Cambridge: Cambridge University Press.

[3] Estlund, D.M., Waldron, J., Grofman, B., and Feld, S. 1989. Democratic theory and the public interest: Condorcet and Rousseau revisited. American Political Science Review 83(4): 1317–1340.

[4] Estlund, D.M. 1994. Opinion leaders, independence, and Condorcet's jury theorem. Theory and Decision 36(2): 131–162.

[5] Goodin, R.E., and Spiekermann, K. 2018. An epistemic theory of democracy. New York: Oxford University Press.

[6] Grofman, B. 1975. A comment on 'democratic theory: A preliminary mathematical model' Public Choice 21(1): 99–103.

[7] Grofman, B., and Feld, S. 1988. Rouseau's general will: a Condorcetian perspective. American Political Science Review 82(2): 567–576.

[8] Rawls, J. 1971. A theory of justice. Cambridge, MA: Harvard University Press.

[9] Wolfram Research, Inc. 2019. Mathematica, Version 12.0, Champaign, IL (2019).

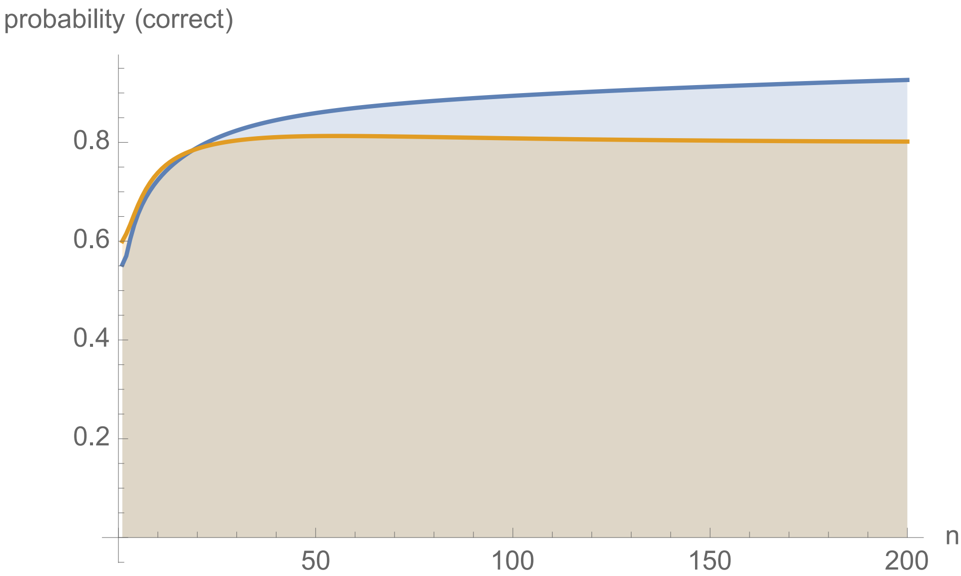

Figure 1: comparison between a good leader and a bad leader.

Note. The blue line represents a bad leader ($q = 0.6$ & $r = 0.05$), while the orange line represents a good leader ($q = 0.8$ & $r = 0.20$). The CJT holds for the former, but fails for the latter. The number of voters is $2n - 1$.

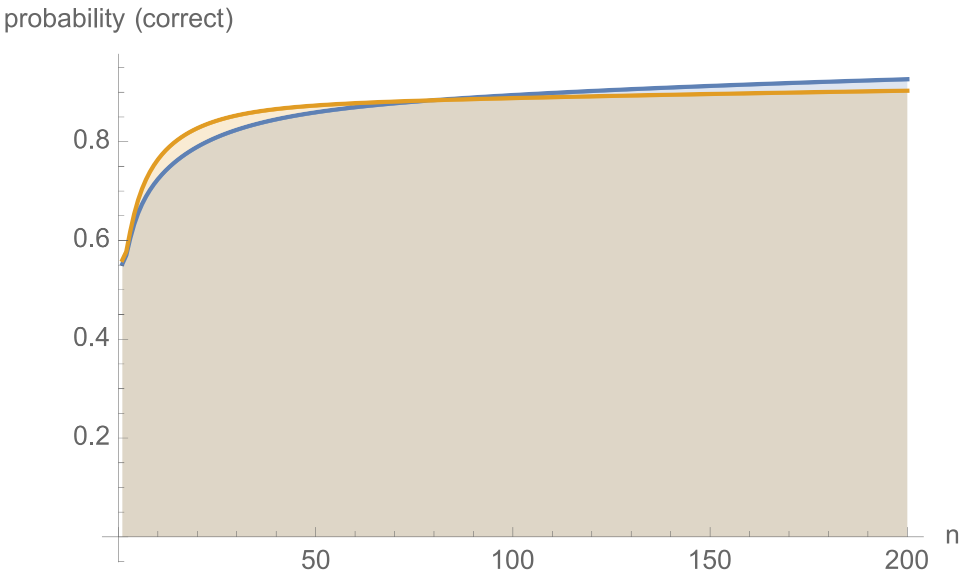

Figure 2: the case with two bad leaders.

Note. The blue line represents a worse leader ($q = 0.6$ & $r = 0.05$), while the orange line represents a less bad leader ($q = 0.7$ & $r = 0.07$). The CJT holds for both. The number of voters is $2n - 1$.

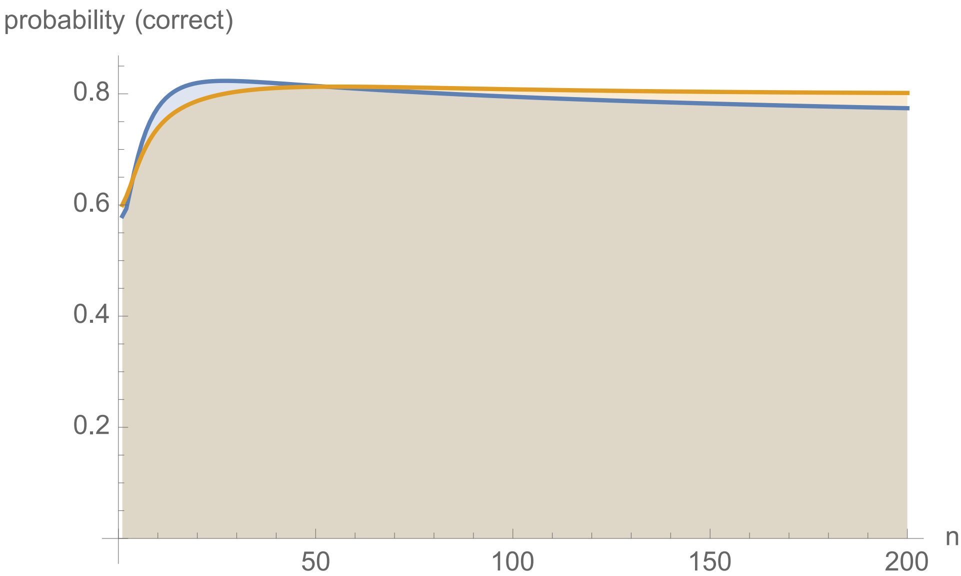

Figure 3: the case with two good leaders.

Note. The blue line represents a less good leader ($q = 0.75$ & $r = 0.15$), while the orange line represents a better leader ($q = 0.8$ & $r = 0.20$). The CJT holds for both. The number of voters is $2n - 1$.