Annex for the scientific article

How to convince in a televised debate: the application of machine learning to analyze why viewers changed their winner perception during the 2021 German chancellor discussion

published in Humanities and Social Sciences Communications

Felix Ettensperger*,

Thomas Waldvogel,

Uwe Wagschal,

Samuel Weishaupt

Felix Ettensperger, Albert-Ludwigs-University, Department of Political Science, Werthmannstraße 12, 79098 Freiburg, felix.ettensperger@politik.uni-freiburg.de.

Thomas Waldvogel, Albert-Ludwigs-University, Department of Political Science, Werthmannstraße 12, 79098 Freiburg, thomas.waldvogel@politik.uni-freiburg.de.

Uwe Wagschal, Albert-Ludwigs-University, Department of Political Science, Werthmannstraße 12, 79098 Freiburg, uwe.wagschal@politik.uni-freiburg.de.

Samuel Weishaupt, Albert-Ludwigs-University, Freiburg, Institute of Computer Science, Georges Köhler Allee, Building 51, 79110 Freiburg, weishaus@informatik.uni-freiburg.de.

***handling author**

1. Handling online streamer data

Including viewers using digital content streaming is a contentious topic in Real-Time-Response Research. Streamers can individually pause streams or experience transfer delays, which might distort response patterns. Often, streamers are getting excluded beforehand to avoid dealing with these distortions. This comes at a heavy price, because larger and larger proportions of data each year thus become unusable for our prediction models. In our sample, only 2.630 participants of 4.613 otherwise valid measurements (only ~57% of the valid sample) responded that they were using a normal Cable or Satellite TV connection to watch the debate. Approximately 40 % of all participants watched the debate via streaming devices.

We argue that our applied machine learning approach can solve parts of this uncertainty problem created by streaming: By measuring the overall performance of the RF prediction, we can evaluate if the models including streaming data are performing better or worse in predicting win perception change, compared to models build only on TV viewer data. Under high noise situations (if e.g. a lot of streamers would be providing distorted, misleading or wrongly coded data) we would expect our full models to decrease in performance. On the other hand, if our model keeps improving and can be trained better while including all cases, we can assume that including these streamers is providing substantially more valid information than noise. Additionally, we can check the variable importance measures of all variables and can evaluate if meaningful distortions become visible in the VIMs patterns (Table A3 and A4).

To check streaming data quality, we completely excluded all streamer data from the sample and evaluate the TV watcher results separately in Appendix section 2, Table A1, A2, figures A1 and A2).

Furthermore, online streamers are on average more progressive, younger and higher educated. Thus, our TV viewer sample also serves as a robustness check using a more representative sample.

2. RF Model Results of exclusively TV viewers (streaming participants excluded)

Table A1: RF Model predicting change to Baerbock.

Sample size: 2630
Frequency of class labels: 2033, 597
Number of trees: 1250
Forest terminal node size: 1
Average no. of terminal nodes: 155.0392
No. of variables tried at each split: 250
Total no. of variables: 314
Resampling used to grow trees: swor
Resample size used to grow trees: 1662
Analysis: RF-C
Family: class
Splitting rule: gini *random*
Number of random split points: 30
Normalized brier score: 39.36
AUC: 90.91
Error rate: 0.14, 0.07, 0.36
##

Confusion matrix:
##
predicted
observed 0 1 class.error
0 1885 148 0.0728
1 216 381 0.3618
##
Overall error rate: 13.88%

Figure A1 – ROCAUC of Baerbock results without streamers.

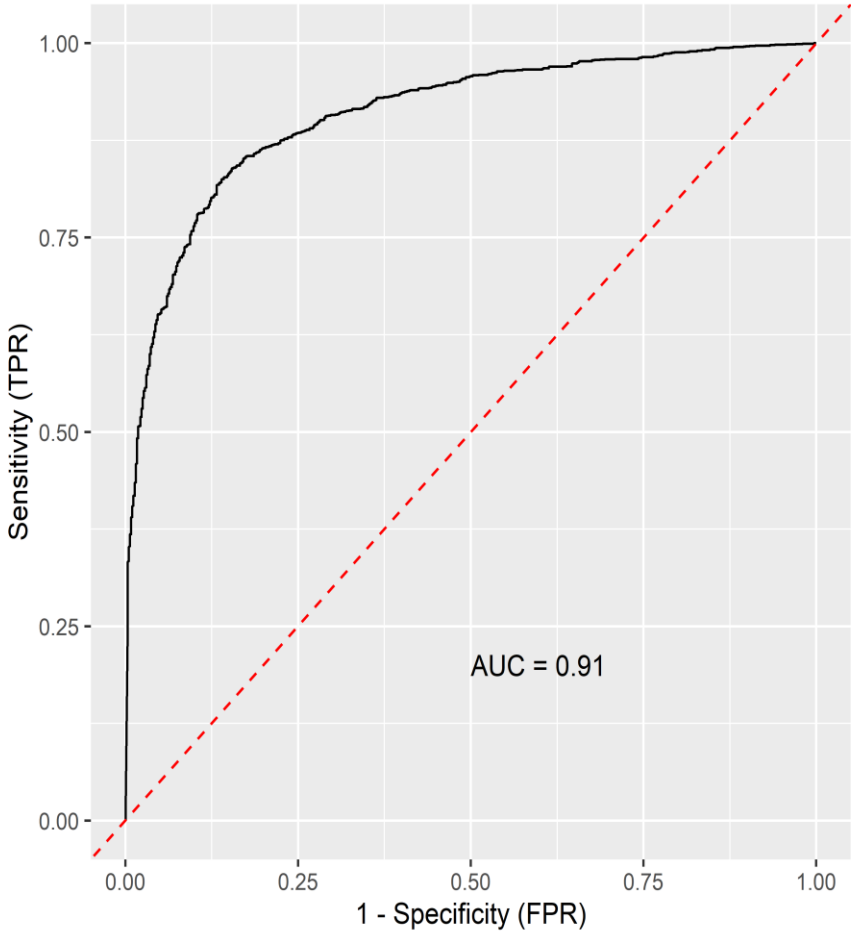
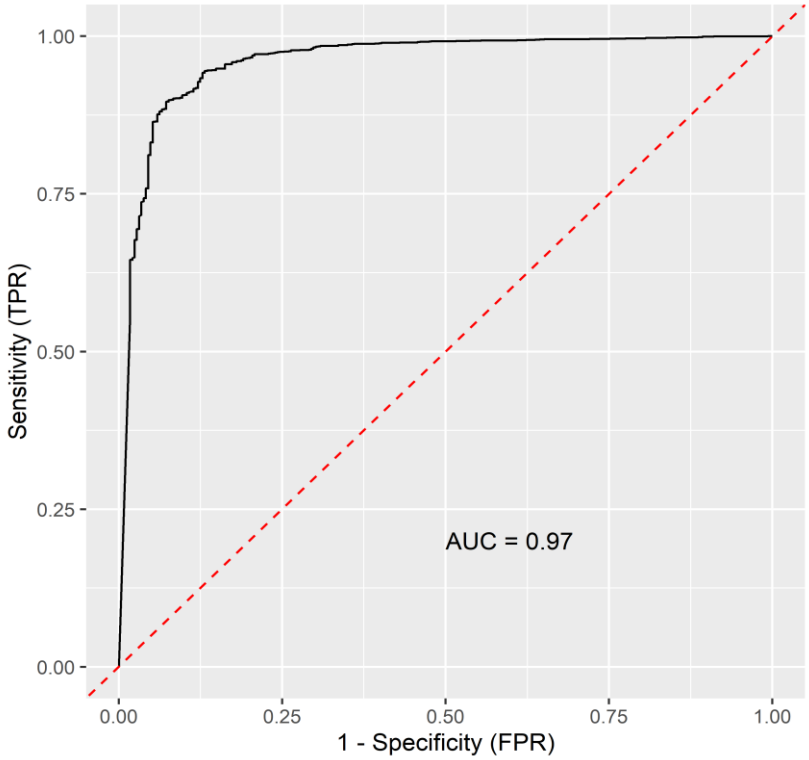


Table A2: RF Model predicting change to Laschet.

Sample size: 2630
Frequency of class labels: 2341, 289
Number of trees: 1250
Forest terminal node size: 1
Average no. of terminal nodes: 72.8616
No. of variables tried at each split: 250
Total no. of variables: 314
Resampling used to grow trees: swor
Resample size used to grow trees: 1662
Analysis: RF-C
Family: class
Splitting rule: gini *random*
Number of random split points: 30
Normalized brier score: 15.38
AUC: 96.03
Error rate: 0.05, 0.03, 0.24
##
Confusion matrix:
##

predicted
observed 0 1 class.error
0 2281 60 0.0256
1 70 219 0.2422
##
Overall error rate: 4.94%

Figure A2 – ROCAUC of Laschet results without streamers.



Evaluating both iterations of predicting change to Baerbock and change to Laschet without streaming viewers provides us with valuable feedback for our research and for RTR research in general:

Both models using TV-only data decrease in overall forecasting quality.

In the Baerbock-Model we experience a drop of overall prediction quality from 87.63 % to 86.12%. The increase in the class error from 26% to 36% is particularly pronounced in the prediction of true positive cases of switching to Baerbock. In the Laschet-Model we experience a drop of prediction quality from 95.99 % to 95.06 %. The class.error in correctly predicting positives is however marginally better in the TV-only model, with 24.2 % versus 24.9 % from the full model.

We can deduct that including streamers from the data will provide us overall with more useful information to better train the algorithm. For the purpose of this research, this means that our variable importance measures are more reliable with this additional data included.

Especially for evaluating change to Baerbock, leaving out streamers would compromise our models and reduce forecasting quality significantly. This is in line with our observation that viewer susceptible to change to Baerbock were more likely to be between age 20-39 (figure 3, main text) and thus more likely to use streaming technology overall.

For viewers that changed their position to Laschet, however, the difference between both iterations is very close. But the full data does perform reliably better in predicting true negatives and only slightly worse in true positives. This candidate dependent effect of excluding streamers is an interesting

finding by itself. We see that excluding streamers introduces an age bias into our sample and that not all candidates are affected equally by this intervention.

In the next tables, we present the top 20 variables of feature importance for the TV-only models.

3. RF Model Feature importance of TV Viewers only

Table A3: Top-20 feature importance of change to Bearbock (tv-only).

	all	0	1
pre.victor	0.1390533	0.2906382	0.6586165
pre.chancellor.ab	0.0454015	0.0942052	0.2174261
pre.ab.comp	0.0056701	-0.0003316	0.0683350
pre.al.comp	0.0016317	-0.0010055	0.0227662
pre.chancellor.al	0.0051533	0.0121300	0.0197793
pre.ab.gred	0.0028424	0.0047814	0.0174116
pre.os.gred	0.0024119	0.0041289	0.0145339
V157	0.0020991	0.0032946	0.0136597
V208	0.0050427	0.0136275	0.0133683
V288	0.0010019	0.0006525	0.0096529
pre.al.gred	0.0012725	0.0016045	0.0096164
pre.ab.lead	0.0023880	0.0055088	0.0095436
V204	0.0010081	0.0011124	0.0081594
V207	0.0022462	0.0056264	0.0074673
pre.chancellor.os	0.0040813	0.0122156	0.0067752
V130	0.0009588	0.0014013	0.0065931
V282	0.0008853	0.0012408	0.0062653
V261	0.0004306	-0.0001070	0.0054639
pre.ab.symp	0.0020867	0.0057120	0.0052818
pre.pol.interest	0.0000865	-0.0009841	0.0043711

Table A4: Top-20 feature importance of change to Laschet (tv-only).

	all	0	1
pre.chancellor.al	0.0547001	0.0981136	0.5444088
pre.victor	0.1302926	0.3543218	0.3195721
V288	0.0137998	0.0304689	0.0910483
pre.al.gred	0.0029506	-0.0014770	0.0842009
pre.chancellor.ab	0.0030495	0.0031677	0.0489855
V157	0.0050192	0.0091964	0.0483835
V42	0.0035675	0.0059637	0.0390530
V191	0.0009249	0.0003530	0.0197898
V273	0.0023730	0.0054342	0.0140711

	all	0	1
pre.ab.comp	0.0003875	-0.0004552	0.0131681
pre.al.comp	0.0008300	0.0009104	0.0129424
pre.ab.gred	0.0008086	0.0008453	0.0129424
V153	0.0003073	-0.0004830	0.0114375
V17	0.0004242	-0.0001208	0.0113622
V262	0.0003688	0.0000279	0.0088038
pre.os.gred	0.0019458	0.0047933	0.0088038
pre.os.comp	0.0007288	0.0012726	0.0075247
pre.ab.symp	0.0002825	0.0000093	0.0068474
V201	0.0000399	-0.0006131	0.0059445
V96	0.0001323	-0.0002787	0.0054930

Evaluating the feature importance values of the TV-only models in comparison to the main results, we conclude that the relevant top 20 variables are mostly identical to the full iterations with 4,613 participants. We do find changes in the ranking of individual variables. But we would expect a certain amount of rank shuffling of variables due to the change in average age in both observational groups.

We do not find evidence in our main results for speaking sections that have been systematically distorted due to lag or streaming pauses. We would typically expect high feature importance sections to systematically radiate into their surrounding temporal neighborhood, finding e.g. high values for V204 and V205 if V203 was an important moment and viewers are giving delayed feedback.

As we are only interpreting the most important features for change, this radiation appears to have not affected our evaluation.

Streaming certainly affects the validity of RTR results in individual cases, and we urge researchers to use tests and precautions to guard against these effects. Nonetheless, we cannot find systematic distortions between our tv-only and the complete sample that would merit the exclusion of so many valuable data points that improve our overall models.

4. Potential Variable Contamination

We use pre-debate winner estimation as a learning variable. An algorithm, like a human expert, can only reliably estimate if the debate induces a change of winner estimation for an individual viewer if the previous position of the viewer is known. If the pre-debate winner estimation would be unknown, the algorithm, like a human expert, would not be able to reliably tell if the RTR response pattern and the pre-debate winner response pattern are consistent or conflicting with each other. But, if pre-debate winner choice and the RTR agreement pattern are highly incompatible, the viewer is more likely to reconsider and change his pre-debate choices. We tested this in various iterations, and in different models without pre-debate winner as learning feature, the algorithm does perform poorly.

In regression-based settings including this variable could raise questions about the validity of the test due to the contamination of the dependent variable (the dependent variable is a combined result of the pre-debate and post-debate evaluation). This is usually referred to as dependent-variable contamination. In our machine-learning setting, this does not constitute a problem, for two main reasons:

First, we are not interested in how much more this pre-debate variable affects the prediction quality in relation to the other variables. It is sufficient to know that this variable is crucial for allowing good

predictions. We do not evaluate the difference between this potentially contaminated variable compared to the others, so we do not need to consider the effect of containment on this individual variable in the analysis.

Secondly, the way machine learning features are included, and their importance is measured fundamentally different compared to the logic of regressions, where a unified mathematical model is constructed from all variables.

Variable importance in ML is measured using an Out-of-bag approach (Hastie et al. 2009: 593) by stepwise exclusion of features. The algorithm is running different iterations of the model excluding each individual variable and evaluating the difference in predictive accuracy of each incomplete model compared to the complete model. Including pre-debate winner estimation thus cannot contaminate the other variables' importance measures in the same way this containment could distort model results in a regression model. We thus can assume that contamination between the pre-debate winner variable and the winner-change variable is not an issue for the interpretation of all the remaining features and their relative importance compared to the other elements in the model.

5. Transcript of speaking moments

The following list of speaking moments is presented in the German original version and an English translation generated by a second-generation machine-learning based translation software. The English translation was further checked by an English literature expert to make sure that all translations are as truthfully to the original version as possible.

Table A5 - Speaking sections of interest.

Speaking section	original debate transcript - German	English language translation
V 204	"Sie haben gefragt, was ist das mit dem Gesetz der Bundesregierung nachdem das Bundesverfassungsgericht ja gesagt hat: "Liebe Große Koalition, so geht das nicht weiter. Ihr gefährdet die Freiheit unserer Kinder. Haben Sie gesagt: OK, wir haben verstanden. Aber was war die einzige Maßnahme, die Sie gemacht haben? Sie haben für in 25 Jahren beschlossen, dann sollten wir beim Klimaschutz weiter sein, aber für hier und heute keine Antwort gegeben." [AB]	**You asked, what is it with the law of the Federal Government after the Federal Constitutional Court has said yes: "Dear Grand Coalition, this can't go on. You are endangering the freedom of our children. You did say: OK, we understand. But what was the only measure you took? You decided that in 25 years we should be further advanced in climate protection, but gave no answer for here and now.** [AB]
V 261	"Fakt ist, dass wir heute ein modernes Einwanderungsgesetz haben. Insofern habe ich gar nicht verstanden, weshalb Sie gesagt haben, wir seien dagegen. Wir haben es. Und wir müssen um qualifizierte Zuwanderung in der Welt werben. Dazu brauchen wir ein neues Klima das signalisiert, wir laden jeden ein, der einen Beitrag leisten kann, nach Deutschland zu kommen. Wir sind gegen eine Einwanderung in die	**The fact is that today we have a modern immigration law. In that respect, I didn't understand at all why you said we were against it. We have it. | And we have to solicit qualified immigration in the world. To do that, we need a new climate that signals that we invite everyone who can make a contribution to come to Germany. We are against immigration into the social systems. Instead, we need to make qualified immigration to**

	Sozialsysteme. Sondern es geht darum, qualifizierte Einwanderung nach Deutschland besser möglich zu machen als das heute der Fall ist. [AL]	**Germany more possible than it is today.** [AL]
V157	"Und das ist die eigentlich große Leistung. Nicht mit Verboten und diesem und jenem. Sondern Tempo machen bei Planung und Genehmigung. Die Trassen schneller bauen um den regenerativen Strom vom Norden in den Süden und in die Industrieregionen zu bringen und gleichzeitig unsere Wirtschaft, die Stahlindustrie, die Zementindustrie, die Chemieindustrie umzugestalten, damit sie klimaneutral wirtschaften kann. Und das muss man können. Das ist keine Frage von Sprüchen, sondern wir man Transformation in unserem Land leistet." [AL]	**And that is actually the great achievement. Not with bans and this and that. But speeding up the planning and approval process. Building the lines faster to bring renewable electricity from the north to the south and to the industrial regions, and at the same time transforming our economy, the steel industry, the cement industry, the chemical industry, so that they can operate in a climate-neutral way. And you have to be able to do that. This is not a question of slogans, but of how to achieve transformation in our country.** [AL]
V207	"Und Sie kommen ja dann immer mit ihren Schlagworten von Verboten. Jedes Verbot ist auch ein Innovationstreiber. Das sehen wir doch jetzt bei den deutschen Automobilherstellern. Wo ich nicht erst im Wahlkampf so wie Sie [AL:"Die tun es doch auch ohne Verbot"] vorbeigeschaut habe. Sondern in den letzten 8 Jahren, in denen ich Wirtschafts- und Klimapolitik mache, ist ganz klar gewesen, die Zukunft im Individualverkehr ist Elektromotor. Und deswegen braucht es vor 2030, muss dafür gesorgt werden, dass dann nur noch emissionsfreie Fahrzeuge zugelassen werden." [AB]	**"And then you always come up with your slogans about bans. Every ban is also an innovation driver. We can see that now with the German car manufacturers. Where I was not only in the election campaign like you. [AL: "They are doing it even without a ban"]. But in the last 8 years, in which I have been making economic and climate policy, it has been very clear that the future in individual transport is the electric motor. And therefore, before 2030, it must be ensured that only emission-free vehicles will then be registered." [AB]**
V137	"Aber die Frage, was will ich anders machen: Zukunftsthemen in das Bundeskanzlerinnenamt integrieren, dass es da eine zentrale Koordinierung gibt. Da braucht man nicht. Da braucht man kein eigenes Ministerium für, sondern wir brauchen ja Digitalisierung in allen Bereichen, der Justiz zum Beispiel. Da konkreter Vorschlag von meiner Partei, dass wir ein Bund-Länder-Programm haben, um zur Digitalisierung der Justiz. Wir erleben da, dass wir viele Straftaten überhaupt nicht vernünftig bundesländerübergreifend erfassen können, weil alles, wie Sie gesagt haben Frau Illner, per Fax hin und her geht." [AB]	**But the question, what do I want to do differently? Integrate the future issues into the Chancellor's Office, that there is central coordination there. You don't need… You don't need a separate ministry for that, but we do need digitization in all areas, the justice system, for example. My party has made a concrete proposal that we have a federal-state program to digitize the judiciary. We're seeing that we can't even record many crimes sensibly across the federal states because, as you said, Ms. Illner, everything goes back and forth by fax.** [AB]

V288	„Liebe Zuschauerinnen und Zuschauer, wir drei haben jetzt 90 Minuten diskutiert, aber es geht nicht um uns. Es geht um sie, es geht um Vertrauen. Ich vertraue unserem Land, unseren Arbeitnehmerinnen und Arbeitnehmern, den Unternehmen, den Gewerkschaften, den Forscherinnen und Wissenschaftlern, dass sie unser Land stark machen können." [AL]	**"Dear viewers, the three of us have now discussed for 90 minutes, but it's not about us. It is about you, it is about trust. I trust our country, our workers, the companies, the unions, the researchers, that they can make our country strong."** [AL]
V153	"Fakt ist, seit den 90er Jahren wissen wir von Weltklimaereignissen. Wir wissen seit Monaten, von Fridays for Future die auf der Straße sind, dass etwas passieren muss. Wir sehen Unwetterkatastrophen überall in der Welt. Wir sehen Brände und merken, wie das Klima sich verändert. Und deshalb ist nicht erst in dem Moment, wo wir auch selbst mal erwischt sind, der Zeitpunkt für eine Politikänderung da. Die findet statt. Sie findet übrigens in der falschen Priorität statt. Wenn das Thema so wichtig ist, wie ich es halte, war die Reihenfolge in Deutschland falsch. Zuerst aus der Kernenergie auszusteigen und dann aus der Kohle. Das ist jetzt Geschichte. Aber der Weg war falsch." [AL]	**"The fact is, we've known about world climate events since the 90s. We've known for months, from Fridays for Future being on the streets, that something has to happen. We see severe weather disasters all over the world. We see fires and we notice how the climate is changing. And that's why it's not just when we're caught out ourselves that the time for policy change is here. It is happening. It's taking place in the wrong priority, by the way. If the issue is as important as I think it is, the order in Germany was wrong. Getting out of nuclear power first, and then out of coal. That's history now. But the path was wrong."** [AL]
V42	"Und Sie sagen, wir haben immer mehr Mitarbeiter. Aber es wird immer weniger aufgeklärt. Und es wird so wenig bei der Geldwäsche aufgeklärt, dass ein Staatsanwalt sagt: Ich brauch hier weitere Informationen. Und dann ein Richter verfügt, dass eine solche Untersuchung stattfinden kann. Und deshalb Herr Scholz war das unangemessen, dass Sie abfällig über die Justiz geredet haben. Das ist unangemessen. Das machen Populisten in anderen Ländern. Wenn die kommen, müssen Sie sagen: Hier ich lege alles offen. Und denen nicht vorschreiben, wie sie zu arbeiten haben. Und wenn es der eine Fall jetzt wäre, wo sie keine Verantwortung übernehmen wollen für den nachgeordneten Bereich den Sie haben. Es ist ja nicht der erste Fall. Bei Wirecard haben Millionen Kleinanleger viel, viel Geld verloren, weil Sie nicht die Aufsicht nicht richtig ausgerichtet haben, die das hätte überprüfen müssen. Dafür tragen Sie Verantwortung. Nicht Sie persönlich, weil Sie da am Schreibtisch sitzen und Aufsicht üben. Aber als Minister haben Sie wie jeder	**"And you say we have more and more employees. But less and less is being cleared up. And so little is being cleared up in money laundering that a prosecutor says, I need more information here. And then a judge decrees that such an investigation can take place. And that's why, Mr. Scholz, it was inappropriate for you to talk disparagingly about the judiciary. That is inappropriate. That's what populists do in other countries. When they come, you have to say: Here, I'll disclose everything. And don't tell them how to work. And if it were the one case now where they don't want to take responsibility for the subordinate area that you have. It's not the first case, after all. In the case of Wirecard, millions of small investors lost a lot of money because you did not properly align the supervisory authority, which should have checked this. You bear responsibility for that. Not you personally, because you are sitting at your desk and exercising supervision. But as a minister, like any other minister, you have to make sure that the supervisory authority that exists**

	andere Fachminister auch dafür zu sorgen, dass die Aufsichtsbehörde, die es gibt, ihren Job macht. Und wenn nicht haben Sie einzugreifen." [AL]	**does its job. And if not, you have to intervene."** [AL]
V17	Nein, wir erleben jetzt ja immer absurdere Theateraufführungen. Vor kurzem wurde Herr Lindner aufgefordert auszuschließen, dass er mit der SPD und den Grünen koaliert. Das hat vielleicht irgendwie deutlich gemacht, dass wir auf dem Boden der Dinge bleiben sollten. Die Bürgerinnen und Bürger entscheiden wie die Wahl ausgeht. Ich werbe für ein sehr sehr starkes Mandat für die SPD. [OS]	**No, we are now witnessing increasingly absurd theatrical performances. Recently, Mr. Lindner was asked to rule out forming a coalition with the SPD and the Greens. That has perhaps somehow made it clear that we should keep our feet on the ground. The citizens decide how the election turns out. I am campaigning for a very very strong mandate for the SPD.** [OS]
V244	"Es ist nicht unbezahlbar. Das ist aus meiner Sicht eine Mär. Und deshalb möchte ich das auch gerne begründen, wenn Sie mir das gestatten. Erstens, wir in den 90er viele Expertinnen und Experten gehört. Die haben uns erzählt, das wird alles nix werden mit der Rente. Das klappt alles nicht mehr. Sie haben uns für diese Zeit vorhergesagt, dass wir sehr hohe Beiträge zahlen werden. Sie haben uns auch gesagt, dass wir viel weniger Arbeitnehmerinnen und Arbeitnehmer haben werden. Das Gegenteil ist eingetreten. Wir zahlen heute geringere Beiträge als zu Zeit von Helmut Kohl. Und wir sind jetzt in einer Situation, wo wir etwa 6 Millionen Arbeitnehmerinnen und Arbeitnehmer mehr haben, die in die Rentenkasse einzahlen. Und deshalb ist die Antwort für die Zukunft auch, dass wir sicherstellen müssen, dass möglichst viele eine gute Beschäftigung haben. Wenn wir voran kommen würden zum Beispiel bei den Beschäftigungsmöglichkeiten und -chancen von Frauen, dann wäre das ein ganz ganz großer Fortschritt und würde die für die Frauen gut sein und für die Rente. Wenn jemand, der mit 56 seinen Job verliert eine plausible Annahme haben kann, dass er auch wieder einen guten neuen Job findet, dann würde das auch schon viel bringen." [OS]	**"It is not unaffordable. That is a fairy tale from my point of view. And that is why I would like to justify it, if you will allow me. Firstly, we heard many experts in the 90s. They told us that there would be no pension. It won't work any more. They predicted that we would pay very high contributions during this period. They also told us that we would have far fewer workers. The opposite has happened. We are paying lower contributions today than when Helmut Kohl was in power. And we are now in a situation where we have about 6 million more workers paying into the pension fund. And therefore the answer for the future is also that we have to make sure that as many as possible have good employment. If we could make progress, for example, on employment opportunities for women, then that would be a very big step forward and would be good for women and for pensions. If someone who loses their job at 56 can have a plausible assumption that they will also find a good new job again, then that would also already bring a lot." [OS]**
V225	„Ich glaube, dass wir zwei Dinge gleichzeitig anpacken müssen: Das eine ist mehr Wohnungen zu bauen. Wir haben es in Deutschland schon Mal geschafft 800.000 Wohnungen in einem Jahr zu bauen. Das war 1973. Da werden wir es wohl jetzt schaffen, jedes Jahr	**“I believe that we need to tackle two things at the same time: One is to build more apartments. In Germany, we have already managed to build 800,000 apartments in one year. That was in 1973, and now we will probably manage to build 400,000 apartments every year,**

	400.000 Wohnungen zu bauen, das sind 100.000 mehr als jetzt. Und darunter müssen auch 100.000 geförderte Sozialwohnungen sein, sonst werden wir das nicht hinbekommen, dass wir in irgend einer Weise das Wohnungsnotproblem in den Städten und den Orten wo viele hinziehen wollen, lösen können.“ [OS]	**which is 100,000 more than we are building now. And that must include 100,000 subsidized social housing units, otherwise we won't be able to solve the housing shortage problem in any way in the cities and places where many people want to move.” [OS]**
V290	„Liebe Zuschauerinnen und Zuschauer, Sie haben die Wahl. Schaffen wir einen echten Aufbruch oder verharren wir im weiter so.“ [AB]	**“Dear viewers, you have a choice. Do we create a real departure or do we persist in business as usual.” [AB]**

Note: AL = Armin Laschet; AB = Annalena Baerbock; OS = Olaf Scholz

6. RF Iteration predicting victory perception change away from Scholz

Appendix Segment 6 presents the RF Iteration results predicting “change away from Scholz”. This is done to demonstrate, that the setup can be reversed and still generates additional useful information about what influenced participants with an “Olaf Scholz pre-preference” to switch to other candidates.

Table A6 – Shift away from Scholz

Sample size: 4613
Frequency of class labels: 3678, 935
Number of trees: 1250
Forest terminal node size: 1
Average no. of terminal nodes: 162.1496
No. of variables tried at each split: 250
Total no. of variables: 320
Resampling used to grow trees: swor
Resample size used to grow trees: 2915
Analysis: RF-C
Family: class
Splitting rule: gini *random*
Number of random split points: 30
(OOB) Normalized Brier score: 25.37709
(OOB) AUC: 96.11424
(OOB) Error rate: 0.08801214, 0.06606852, 0.17433155
##
Confusion matrix:
##
predicted
observed 0 1 class.error
0 3435 243 0.0661
1 163 772 0.1743
##
Overall (OOB) error rate: 8.801214%

Figure A3 – ROCAUC of Switch away from Scholz.

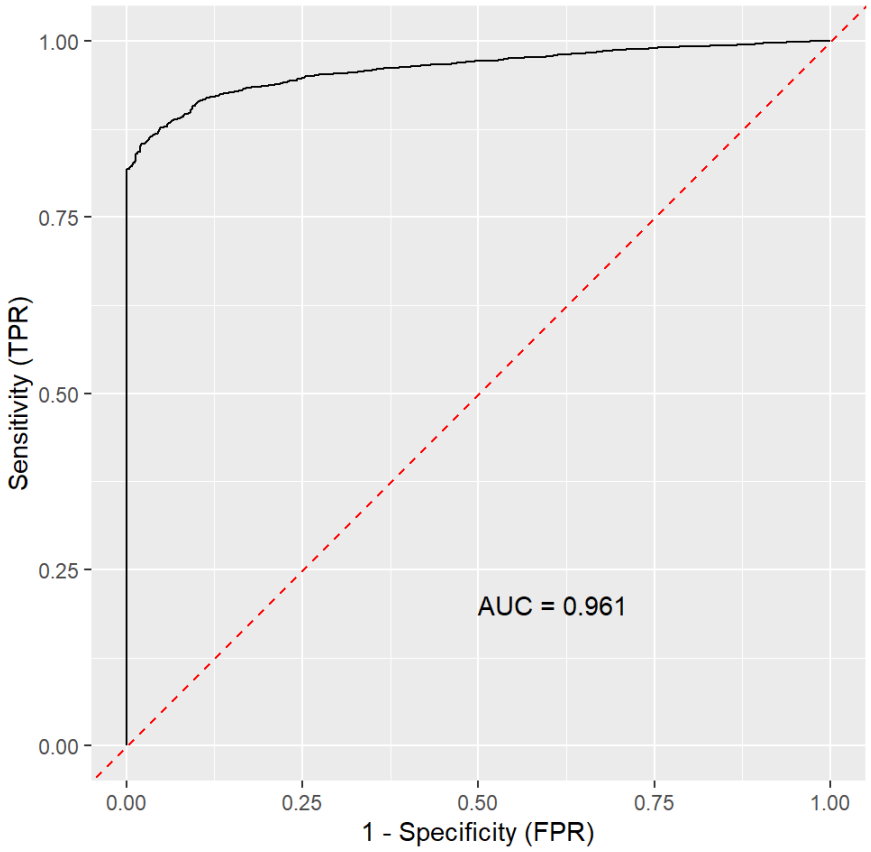


Table A7 – VIMs of Scholz switch away.

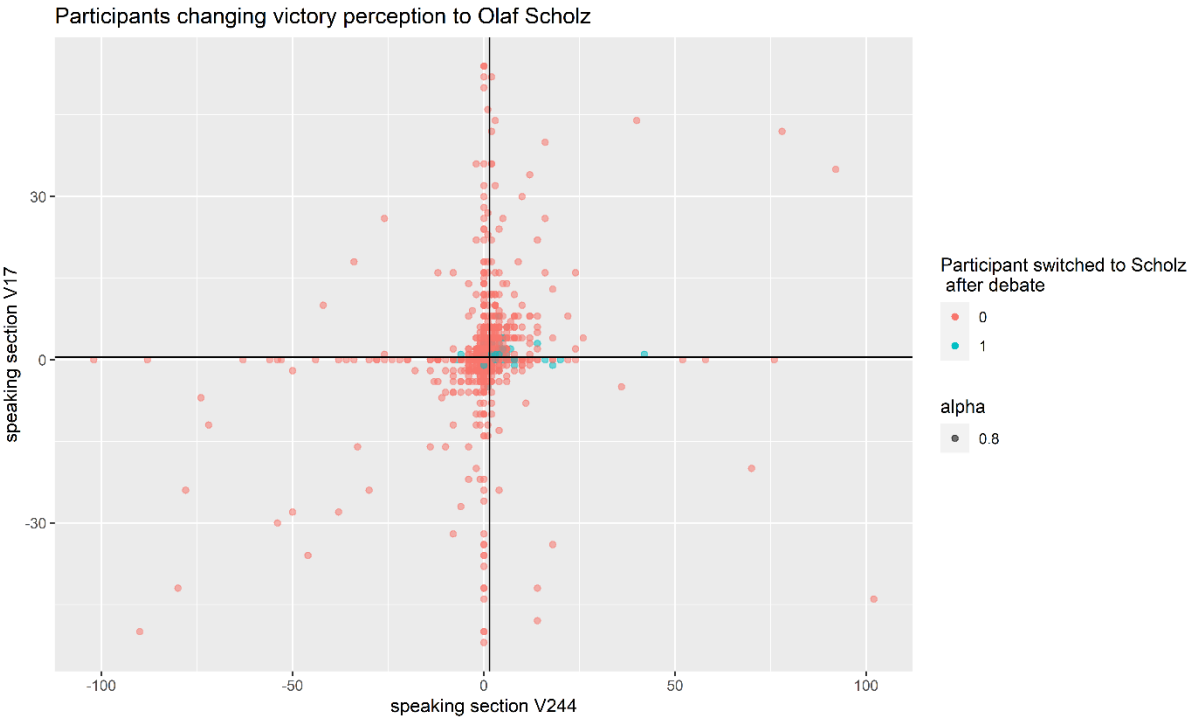
	all	0	1
pre.victor	0.2416128	0.4308750	1.5123066
V207	0.0025811	0.0059362	0.0109080
pre.chancellor.al	0.0005151	-0.0007450	0.0097684
V204	0.0006799	0.0003015	0.0078380
V282	0.0002979	0.0000532	0.0037445
pre.age	0.0001315	-0.0004730	0.0036050
pre.al.comp	0.0005920	0.0013540	0.0025351
V213	0.0003223	0.0004848	0.0023723
V137	0.0004172	0.0008514	0.0021863
V81	0.0005431	0.0013244	0.0020002
V288	0.0003207	0.0006504	0.0016978
pre.ab.symp	0.0003152	0.0006504	0.0016281
V157	0.0001647	0.0001951	0.0014187
V120	0.0002891	0.0006208	0.0013955
V262	0.0002786	0.0006326	0.0012094
V241	0.0000718	-0.0000473	0.0011396
V136	0.0001176	0.0001242	0.0010699
V40	0.0003030	0.0007627	0.0010234
V171	0.0002523	0.0005972	0.0010001
V268	0.0001086	0.0001360	0.0009071

We see that especially the debate moments V207, V204, and V282 determined if participants changed their winner estimation away from Scholz during the debate. But also pre.age and pre.al.comp (Competence evaluation of Laschet) and pre.ab.symp (Sympathy evaluation of Baerbock) affected strongly if participants changed winner expectation from Scholz to another candidate in the course of the debate.

7. Additional Figures and Analysis

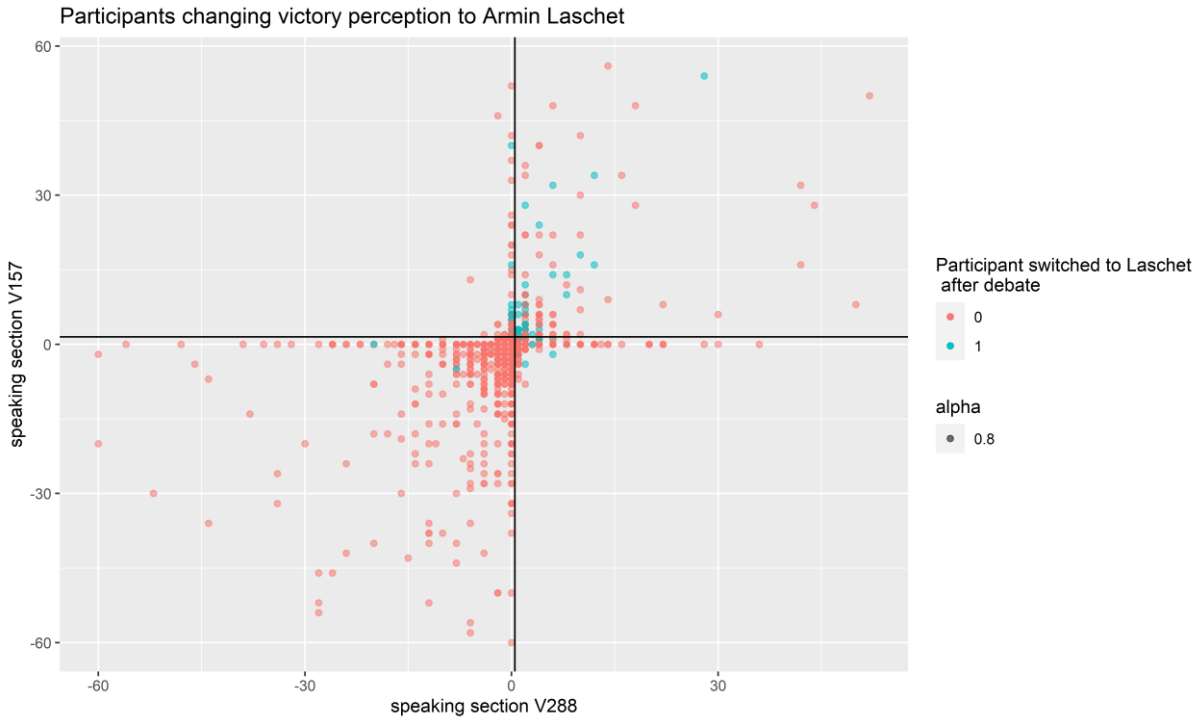
The pathway I is visualized in figure A3, indicating the relationship between V17 and V244 for shift to Scholz. Only few participants and most of them close to the threshold experienced a change (figure A3). In additional tests we found that higher agreement was already corresponding with a pre-debate winner expectation for Scholz, lower agreement did not lead to change. Thus, we see positive cases of change clustering around the threshold.

Figure A3. Pathway I: Combination of V17 and V244



Another relationship we visualized is between the combination of V157 and V288 for change to Laschet in Figure A4. It shows the interaction of a policy related response and a patriotic address. Again, we see a special high number of switching participants close to the thresholds.

Figure A4. Pathway H: Combination of V157 and V288



Most participants with the RTR response pattern V157 >= 0.5 & V288 >= 1.5 are already expecting Laschet to win. Those who don't are mainly expecting Olaf Scholz or a draw in their pre.debate expectation and many of these participants are changing their expectation during the debate.

Table A8 – First 25 cases of Participants with the combination V157 >= 0.5 & V288 >= 1.5

	user_id	pre.victor	victory.change.AL	no.change
1	10	Armin Laschet	0	1
2	19	Olaf Scholz	1	0
3	30	Armin Laschet	0	1
4	35	Armin Laschet	0	1
5	61	Armin Laschet	0	1
6	122	Armin Laschet	0	1
7	154	draw	1	0
8	155	Armin Laschet	0	1
9	163	Armin Laschet	0	1
10	172	Armin Laschet	0	1
11	253	Olaf Scholz	0	0
12	355	draw	0	1
13	368	draw	1	0
14	400	Armin Laschet	0	1
15	454	Olaf Scholz	0	0
16	471	draw	1	0
17	530	draw	1	0
18	549	draw	1	0
19	561	Armin Laschet	0	1

20	570	draw	1	0
21	608	Armin Laschet	0	1
22	652	draw	1	0
23	656	Armin Laschet	0	1
24	673	Olaf Scholz	1	0
25	675	no response	0	0

…

First 25 of 214 participants

8. Robustness checks of RF and decision tree models

8.1 Random Forest robustness checks

8.1.1 Random Forest Models predicting with only pre-debate survey data:

We added these tests to determine if using only pre-debate data generated competitively good results compared to the full model. This would indicate that debate moment inclusion has only a marginal effect on predictive accuracy. In this case we would have to concede that our Hypothesis 2 is not viable.

We observe that pre-debate data in isolation is better than using only RTR data for predictions (compare with table 8.1.2). But the pre-debate data tests perform clearly worse than the full model including pre-debate data and RTR data combined. H1 and H2 in combination seem to be best for forecasts.

Shift to Annalena Baerbock

```
##                         Sample size: 4613
##           Frequency of class labels: 3522, 1091
##                     Number of trees: 750
##           Forest terminal node size: 1
##       Average no. of terminal nodes: 467.7813
## No. of variables tried at each split: 5
##              Total no. of variables: 21
##       Resampling used to grow trees: swor
##     Resample size used to grow trees: 2915
##                             Analysis: RF-C
##                               Family: class
##                     Splitting rule: gini *random*
##       Number of random split points: 10
##        (OOB) Normalized Brier score: 36.3332
##                           (OOB) AUC: 92.91829
##                  (OOB) Error rate: 0.12876653, 0.07410562, 0.30522456
##
## Confusion matrix:
##
##          predicted
##   observed    0   1 class.error
##          0 3262 260      0.0738
##          1  334 757      0.3061
##
##  Overall (OOB) error rate: 12.876653%
```

Shift to Armin Laschet

```
##                         Sample size: 4613
##           Frequency of class labels: 4188, 425
##                     Number of trees: 750
##           Forest terminal node size: 1
##       Average no. of terminal nodes: 230.508
## No. of variables tried at each split: 5
##              Total no. of variables: 21
##       Resampling used to grow trees: swor
##     Resample size used to grow trees: 2915
##                             Analysis: RF-C
##                               Family: class
##                      Splitting rule: gini *random*
##       Number of random split points: 10
##        (OOB) Normalized Brier score: 15.22712
##                           (OOB) AUC: 95.42994
##                   (OOB) Error rate: 0.04682419, 0.01241643, 0.38588235
##
## Confusion matrix:
##
##           predicted
##   observed    0   1 class.error
##          0 4136  52      0.0124
##          1  165 260      0.3882
##
##  Overall (OOB) error rate: 4.682419%
```

8.1.2 Random Forest Models predicting with only RTR data

We added these tests to determine if using only pre-debate data performs competitively good compared to the full model. This would indicate that pre-debate data inclusion has only a marginal effect on predictive accuracy. In this case we would have to concede that our Hypothesis 1 is not viable.

We observe that RTR data in isolation is less suitable for predictions than the pre-debate models (compare with table 8.1.1) and the full models. The RTR data test perform clearly worse than the full model including pre-debate data and RTR data combined. H1 and H2 in combination seem to be best for forecasts.

Shift to Annalena Baerbock

```
##                         Sample size: 4613
##           Frequency of class labels: 3522, 1091
##                     Number of trees: 1250
##           Forest terminal node size: 1
##       Average no. of terminal nodes: 406.8064
## No. of variables tried at each split: 250
##              Total no. of variables: 294
##       Resampling used to grow trees: swor
##     Resample size used to grow trees: 2915
##                             Analysis: RF-C
##                               Family: class
##                      Splitting rule: gini *random*
##       Number of random split points: 30
##        (OOB) Normalized Brier score: 42.40086
##                           (OOB) AUC: 90.59134
##                   (OOB) Error rate: 0.15759809, 0.12322544, 0.26856095
##
## Confusion matrix:
##
##           predicted
##   observed    0   1 class.error
##          0 3089 433      0.1229
##          1  293 798      0.2686
##
##  Overall (OOB) error rate: 15.759809%
```

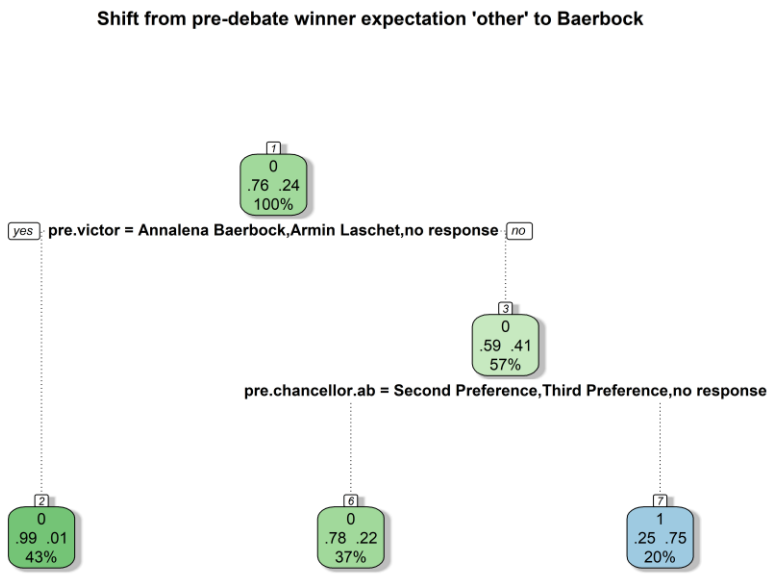
Shift to Armin Laschet

```
##                         Sample size: 4613
##            Frequency of class labels: 4188, 425
##                      Number of trees: 1250
##            Forest terminal node size: 1
##        Average no. of terminal nodes: 236.984
## No. of variables tried at each split: 250
##               Total no. of variables: 294
##        Resampling used to grow trees: swor
##     Resample size used to grow trees: 2915
##                             Analysis: RF-C
##                               Family: class
##                     Splitting rule: gini *random*
##        Number of random split points: 30
##         (OOB) Normalized Brier score: 17.08255
##                              (OOB) AUC: 93.18973
##                     (OOB) Error rate: 0.05441145, 0.02292264, 0.36470588
##
## Confusion matrix:
##
##           predicted
##   observed    0   1 class.error
##          0 4092  96      0.0229
##          1  155 270      0.3647
##
##  Overall (OOB) error rate: 5.441145%
```

8.2 Decision Tree robustness checks

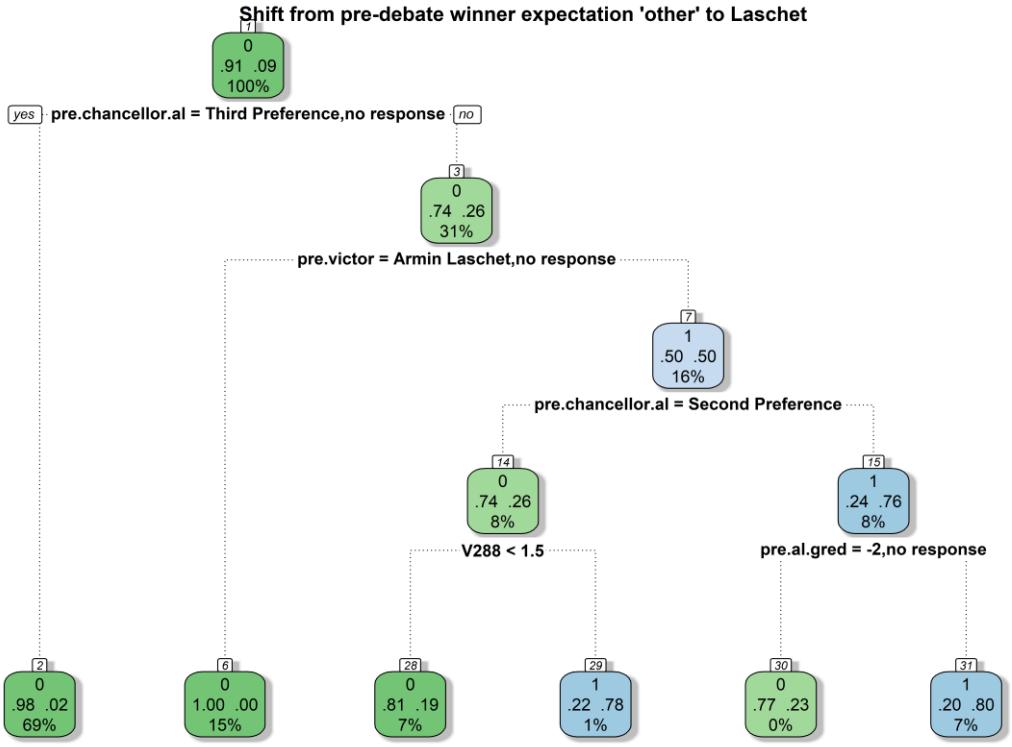
Adding the full model decision trees to the annex helps us to identify viable connections between pre-debate data and RTR response effects.

Decision tree shift to Baerbock: Including pre-debate data:



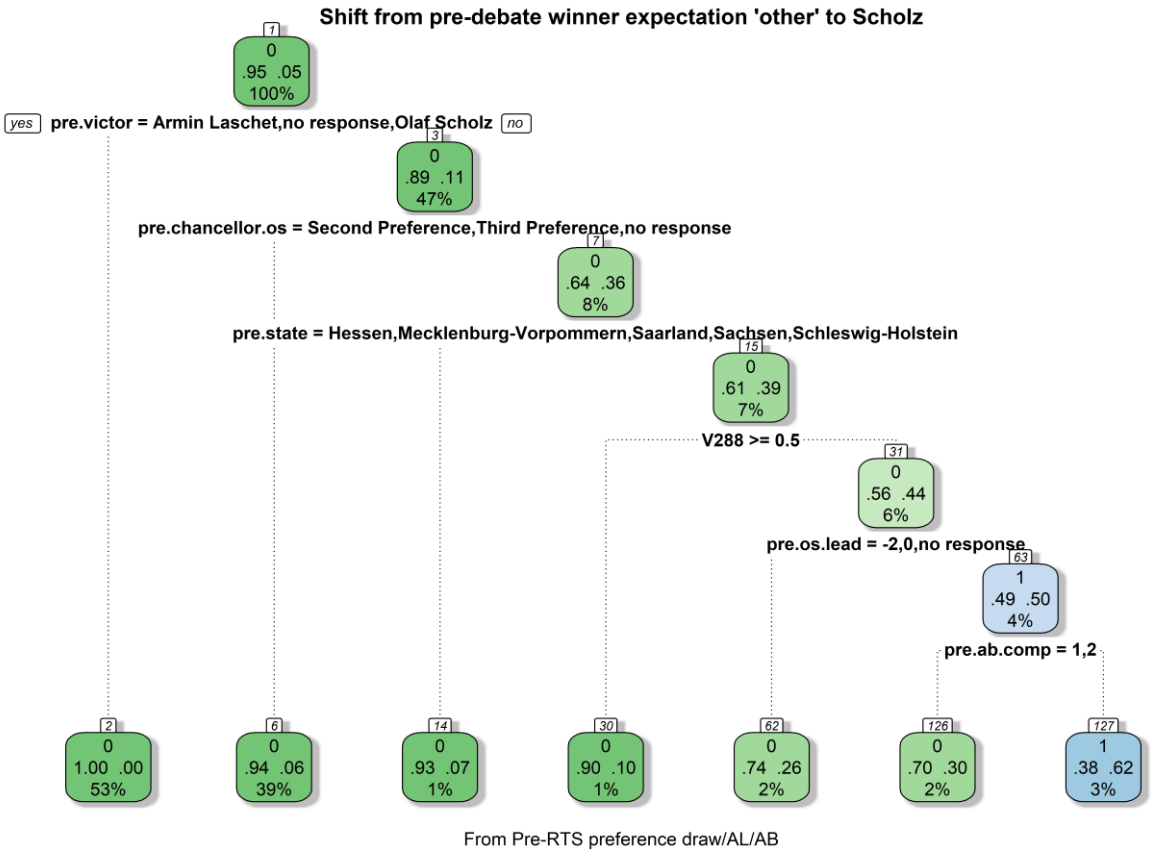
From Pre-RTS preference draw/OS/AL

Decision tree shift to Laschet: Including pre-debate data:



From Pre-RTS preference draw/OS/AB

Decision tree shift to Scholz: Including pre-debate data:



From Pre-RTS preference draw/AL/AB

9. Data Summaries

Relationship of Pre-victor expectation, Party identification and Post-victor evaluation. More detailed summary statistics in our R Script.

9.1 Post-debate Participants evaluation: Armin Laschet has won the debate.

pre.debate expectation	Number of participants
Annalena Baerbock	25
Armin Laschet	638
Olaf Scholz	153
draw	247
No response	24

post.debate evaluation	Number of participants
Annalena Baerbock	0
Armin Laschet	1087
Olaf Scholz	0
draw	0
No response	0

pre.debate party id	Number of participants
CDU/CSU	731
FDP	146
No party id	104
AfD	39
SPD	24
Green Party / B90	21
Die Linke	2
No response	20

9.2 Post-debate Participants evaluation: Olaf Scholz has won the debate.

pre.debate expectation	Number of participants
Annalena Baerbock	89
Armin Laschet	12
Olaf Scholz	740
draw	149
No response	20

post.debate evaluation	Number of participants
Annalena Baerbock	0
Armin Laschet	0
Olaf Scholz	1010
draw	0
No response	0

pre.debate party id	Number of participants
CDU/CSU	92
FDP	44
No party id	115

AfD	13
SPD	434
Green Party / B90	198
Die Linke	84
No response	30

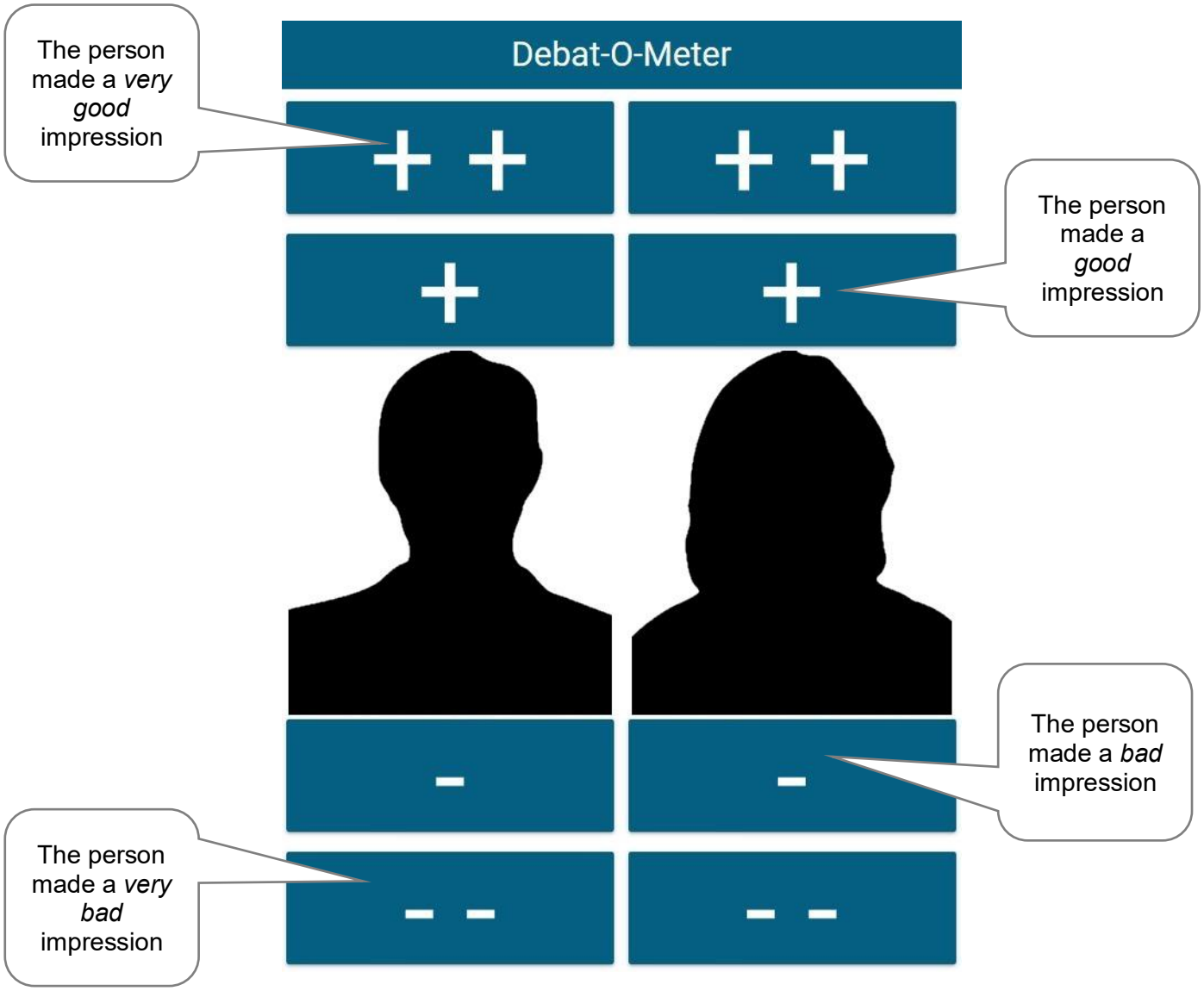
9.3 Participants with post-debate evaluation: Annalena Baerbock has won the debate.

pre.debate expectation	Number of participants
Annalena Baerbock	1045
Armin Laschet	19
Olaf Scholz	660
draw	412
No response	52

post.debate evaluation	Number of participants
Annalena Baerbock	2188
Armin Laschet	0
Olaf Scholz	0
draw	0
No response	0

pre.debate party id	Number of participants
CDU/CSU	110
FDP	41
No party id	249
AfD	1
SPD	259
Green Party / B90	1187
Die Linke	261
No response	80

10. Measurement Instructions to the participants.



Please note the following:

- The Debat-O-Meter records your ratings exact to the second, which means one rating is recorded per second.
- Why you rate a person and what exactly you consider good or bad is left entirely to you.
- If no button is pressed, no data is transmitted, which will be interpreted as a "neutral" rating.

Please do not use the device when the host is speaking.

11. Additional Information regarding the applied algorithms

a) Decision tree algorithm

Decision Trees are constructed by recursively selecting different features of the dataset for each node in the decision tree and evaluating what feature leads to the best and clearest split in the data in the target variable following this mathematical procedure:

Suppose that we have a scalar outcome, Y , and a p-vector of explanatory variables, X.

Assume $Y \epsilon K = \{1, 2, \ldots, k\}$*. The subsets created by the splits are called nodes. The subsets which are not split are called terminal nodes.*

Each terminal nodes gets assigned to one of the classes. So, if we had 3 classes we could get $A_1 = X_5 \cup X_9$, $A_2 = X_6$ *and* $A_3 = X_7 \cup X_8$*. If we are using the data we assign the class most frequently found in that subset of X.*

A classification tree partitions the X-space and provides a predicted value, according to arg max_s $Pr(Y = s \mid X \in A_k)$ *in each region*

(Breiman et al. 1984: 255-257).

For more details and a full technical explanation on the algorithm of decision trees please consider the literature of Breiman et al. 1984, James et al. 2013: 303, Hastie et al. 2009: 305.

b) Random Forest regression tree / classification tree

Based on decision trees the RF Algorithm combines randomized tree ensembles to create even better forecasts. The algorithm applied is defined as:

1. *For b = 1 to B:*

    (a) Draw a bootstrap sample $Z*$ *of size N from the training data.*

    (b) Grow a random-forest tree T_b *to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size* n_{min} *is reached.*

i. Select m variables at random from the p variables.
ii. Pick the best variable/split-point among the m.
iii. Split the node into two daughter nodes.

2. Output the ensemble of trees $\{Tb\}_1^B$.

To make a prediction at a new point x:

Regression: $f_{rf}^{B}(x) = \frac{1}{B} \Sigma_{b=1}^{B} T_b(x)$

Classification: Let $C_b(x)$ *be the class prediction of the bth random-forest tree. Then* $C_{rf}^{B}(x) = majority\ vote \{ C_b(x) \}_1^B$

(Hastie et al. 2009: 588)

12. Additional Comments from the authors

„Die Autoren möchten an dieser Stelle das umfassende Grundlagenwerk von Maier und Faas (2019) ausdrücklich als einen wichtigen Beitrag zur systematischen Aufarbeitung des Forschungsstandes zur politischen Debattenkommunikation im deutschsprachigen Raum würdigen. Der Band bietet eine fundierte und gut strukturierte Übersicht über zentrale theoretische Konzepte, empirische Befunde und methodische Zugänge der Debattenforschung mit einem Fokus auf den deutschen Kontext. Er war daher auch für die Erstellung des vorliegenden Beitrags in einer initialen Recherchestufe von orientierender Bedeutung: für unsere Darstellung des Forschungsstands wurde an verschiedenen Stellen die von Maier & Faas (2019) verwendete Strukturierung und von ihnen vorgenommene Zusammenfassungen übernommen. Auch wenn im vorliegenden Artikel der Fokus auf international anschlussfähige, englischsprachige Literatur gelegt wurde, um der internationalen Ausrichtung des Journals und der Leserschaft gerecht zu werden, sei allen Leser:innen die Lektüre des Studienbuches von Maier und Faas (2019) für einen umfassenden Einblick in die deutschsprachige Debattenforschung empfohlen.“

“The authors would like to take this opportunity to expressly acknowledge the comprehensive work by Maier and Faas (2019) as an important contribution to the systematic review of the state of research on political debate communication in German-speaking countries. The volume offers a well-founded and well-structured overview of central theoretical concepts, empirical findings and methodological approaches to debate research with a focus on the German context. It was therefore also of orienting importance for the preparation of this article in an initial research stage: for our presentation of the state of research we applied a structural design aligned to their previous work (Maier & Faas 2019). Even though the focus in this article was placed on internationally compatible, English-language literature in order to do justice to the international orientation of the journal and its readership, all readers are recommended to read the study book by Maier and Faas (2019) for a comprehensive insight into German-language debate research.”

Maier, J. & Faas, T. (2019). TV-Duelle. Wiesbaden: Springer VS.