

Online-Anhang zu Gschwend, Juhl, und Lehrer (2018) „Die „Sonntagsfrage“, soziale Erwünschtheit und die AfD: Wie alternative Messmethoden der Politikwissenschaft weiterhelfen können“

A1. Test der Randomisierung

Wir teilen Befragte zufällig in drei Gruppen ein, die jeweils einen bestimmten Fragentyp gestellt bekommen. Ein zufällig ausgewähltes Drittel der Befragten wurde der RAT-Befragung zugeordnet, das zweite zufällig ausgewählte Drittel erhielt die Listenexperiment-Fragen mit Liste A als Treatment und Liste B als Kontrollliste und das letzte Drittel der Befragten erhielt ebenfalls die Listenexperiment-Fragen, jedoch mit Liste B als Treatment und Liste A als Kontrollliste. Die zufällige Einteilung soll erreichen, dass die drei Gruppen in allen wesentlichen Merkmalen gleich und somit vergleichbar sind. In diesem Anhang zeigen wir, dass dies zwar nicht perfekt, aber doch in beachtlicher Weise gelingt. Abbildung 5 zeigt die Anzahl der Befragten, die bestimmten Charakteristika zugeordnet werden können. Diese sind das (1) Geschlecht der Befragten, (2) eine etwa fünfjährige Jahresspanne, in der die Befragten geboren wurden, (3) das Datum, an dem die Befragten die Umfrage ausfüllten, (4) ob sie in der „Sonntagsfrage“ eine AfD-Wahlabsicht äußern, (5) ob sie in den alten oder neuen Bundesländern wohnen, (6) welcher ihr höchster Schulabschluss ist und (7) ob sie auch ohne das GIP über einen Internetzugang verfügen.

Die letzte Eigenschaft bedarf einer kurzen Erklärung. Bei der Rekrutierung der GIP-Teilnehmerinnen und -Teilnehmer wurde die gesamte deutsche Bevölkerung zwischen 16 und 75 Jahren zugrunde gelegt und dies unabhängig davon, ob die ausgewählten Personen über einen Internetzugang oder die Computerkenntnisse verfügen, um an einer Online-Befragung teilzunehmen. Befragten, die über mindestens einen dieser Faktoren nicht verfügen, wurden sowohl ein für Computer-Anfänger entwickelter Computer sowie ein Internetzugang bereitgestellt (Blom et al. 2017). Wir vergleichen deshalb zwischen den Gruppen, wie viele Personen vom GIP-Team mit einem Internetanschluss oder Computer ausgestattet wurden.

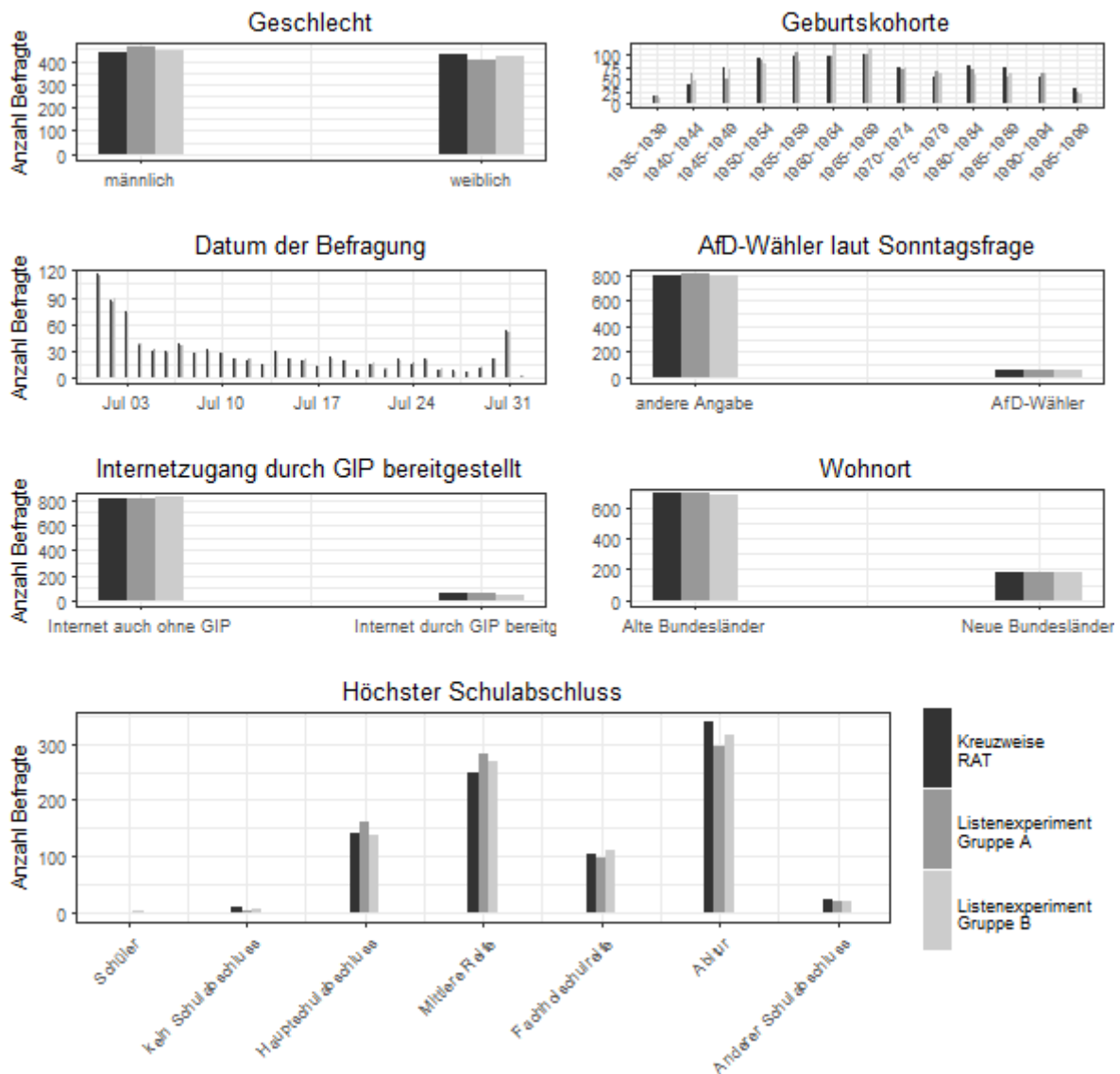


Abb. 5 Verteilung zentraler Charakteristika unter den Befragungsgruppen. *Anmerkungen:* Alle Grafiken sind Histogramme. Listenexperiment (A) = Liste A enthält fünf Aussagen, Listenexperiment (B) = Liste B enthält fünf Aussagen. RAT ist die Randomisierte-Antwort-Technik. *Quelle:* Eigene Berechnungen auf Grundlage der 30. Welle des German Internet Panels.

Die verschiedenen Säulen in Abbildung 5 stellen die verschiedenen Gruppen dar. Sind diese Säulen annähernd gleichhoch, so gleichen sich die Gruppen in der dargestellten Eigenschaft. Auf den ersten Blick ist sichtbar, dass die Gruppen sich tatsächlich stark gleichen (siehe etwa Geschlecht, AfD-Wählerin/Wähler laut „Sonntagsfrage“, Internetzugang durch GIP bereitgestellt und Wohnort). Die Eigenschaften mit vielen Merkmalsausprägungen (Datum der Befragung, Geburtskohorte und höchster Schulabschluss) weichen zwar in Einzelfällen voneinander ab, würden sich aber annähernd wieder ausgleichen, wenn benachbarte Kategorien zusammengefasst würden. Da die Einteilung in die Kategorien zu einem gewissen Grad willkürlich ist (zum Beispiel die Einteilung der Geburtskohorten), sind wir überzeugt, dass die vereinzelt sichtbaren Abweichungen unsere Ergebnisse nicht wesentlich beeinflussen

A2. Verteilung der ersten Ziffer von Hausnummern in Deutschland

Die Verteilung der ersten Ziffern der Hausnummern aller GIP-Befragten haben wir ausgewertet, um die Annahme der Bendforsche Verteilung zu überprüfen. Dies ist ein zulässiger Test, da es sich bei den GIP-Befragten um eine repräsentative Stichprobe der Bevölkerung handelt. Abbildung 6 zeigt, dass unsere Verteilungsannahme gerechtfertigt ist. Circa 70 % der Hausnummern der Befragten beginnen mit Ziffern zwischen 1 und 4.

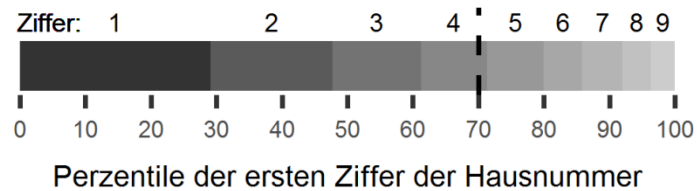


Abb. 6 Verteilung der ersten Ziffer der Hausnummern der GIP-Befragten. *Quelle:* Eigene Darstellung auf Grundlage von Daten, die vom GIP-Team zu Verfügung gestellt wurden.

A3. Analyse der Antwortzeiten in Sekunden

Im Artikel zeigen wir, dass die Weisheit-der-Vielen-Methode auch dann noch am besten abschneidet, wenn man verschiedene Perzentile der schnellsten Antworten ausschließt. Hier zeigen wir, dass dies auch dann passiert, wenn Fälle nicht nach Perzentilen der Beantwortungszeit, sondern nach verstrichener Zeit (in Sekunden) ausgeschlossen werden.

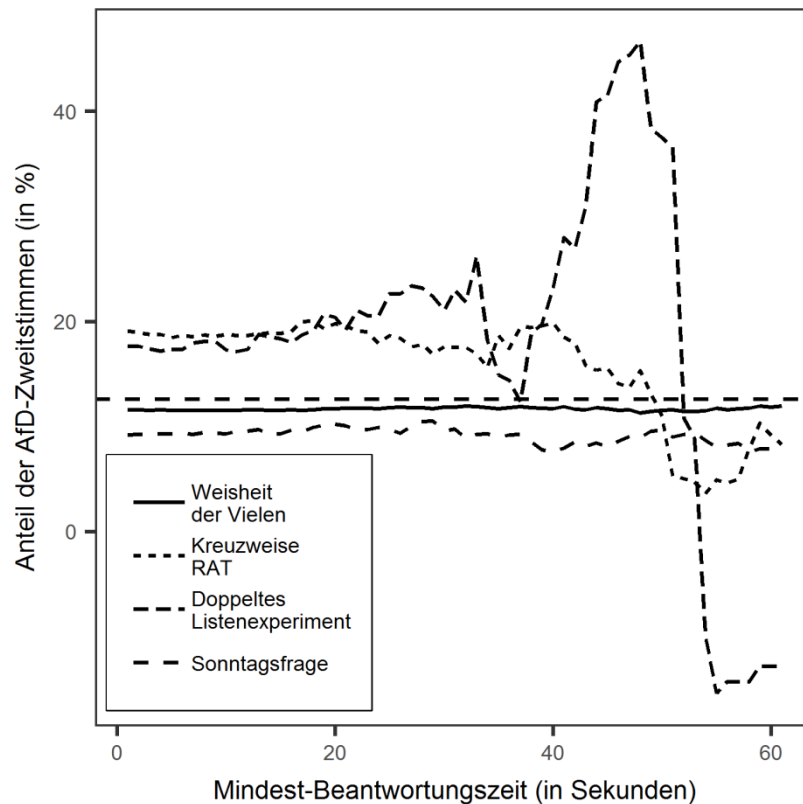


Abb. 7 Erwarteter AfD-Zweitstimmenanteil bei der Bundestagswahl 2017 nach Antwortzeit. *Anmerkungen:* Die schwarzen Linien stellen Punktschätzer dar, die graue Linie das tatsächliche AfD-Zweitstimmenergebnis. RAT ist die Randomisierte-Antwort-Technik. *Quelle:* Eigene Berechnungen auf Grundlage der 30. Welle des German Internet Panels.

Dazu zeigt Abbildung 7, wie sich die Vorhersagen der verschiedenen Methoden entwickeln, wenn man Antworten ignoriert, die in weniger als einer bestimmten Zeit abgegeben wurden (x-Achse). Am linken Ende des Graphen (Mindest-Beantwortungszeit = 0 Sekunden) entsprechen die Vorhersagen den Prognosen aus Abbildung 1. Mit zunehmender Zeit fallen mehr und mehr Beobachtungen aus dem Pool der Fälle heraus, aus dem die 865 Fälle für die Analyse zufällig mit Zurücklegen gezogen werden.

Auch wenn es an einzelnen Stellen im Detail Unterschiede gibt, sind die Ergebnisse und Schlüsse, die wir im zugehörigen Artikel selbst ziehen, nicht davon abhängig, ob man Fälle nach absoluter Zeit oder Perzentilen der Beantwortungszeit ausschließt. In beiden Fällen liegt die Weisheit-der-Vielen-Methode am nächsten am wahren Ergebnis. Zusätzlich unterliegt ihre Vorhersage den wenigsten Schwankungen. Wir sehen darin weitere Evidenz, dass die Weisheit-der-Vielen-Methode eine sinnvolle Ergänzung zur „Sonntagsfrage“ ist.