

Das Debat-O-Meter: ein neues Instrument zur Analyse von TV-Duellen

von Thomas Metz, Uwe Wagschal, Thomas Waldvogel, Marko Bachl,
Linus Feiten und Bernd Becker

I. Einleitung

Live im Fernsehen übertragene Debatten zwischen politischen Spitzenkandidaten (*TV-Duelle*) sind das wohl augenfälligste Kennzeichen einer medial vermittelten Massendemokratie. Als US-Import 2002 erstmals zur Bundestagswahl eingesetzt, hat das Format schnell seinen Platz in deutschen Wahlkämpfen erobert.¹ Wichtigster Grund hierfür ist, dass der Breitenwirkung von TV-Duellen das Potenzial zugeschrieben wird, den Wahlkampf entscheidend zu beeinflussen.²

Die Wirkung von TV-Duellen verbindet sich mit mehreren Aspekten³: Zunächst erreicht kein anderes Wahlkampfeignis derart viele Wähler; so sahen beispielsweise zwischen 14 und 21 Millionen Zuschauer die bisherigen Duelle auf Bundesebene.⁴ Zum zweiten deckt kein anderes Format eine derart breite Palette an Themen ab und erlaubt es Wählern, nicht nur die Kandidaten und ihre Positionen kennen zu lernen, sondern diese auch direkt zu vergleichen. Und drittens sprechen TV-Duelle in besonderem Maße auch politisch weniger interessierte und daher oft nur schwer zu erreichende Wähler an⁵, sodass sie die Beschäftigung mit Politik anstoßen und den Boden für politisches Engagement bereiten können. *Faas* und *Maier* sprechen deshalb zurecht von „Wahlkämpfe[n] im Miniaturformat“.⁶

1 *Maier, J./Faas, T./Maier, M.*: Aufgeholt, aber nicht aufgeschlossen: Ausgewählte Befunde zur Wahrnehmung und Wirkung des TV-Duells 2013 zwischen Angela Merkel und Peer Steinbrück, in: Zeitschrift für Parlamentsfragen, 45/1 (2014), 38–54, hier 40.

2 *Maier, J./Faas, T.*: 'Miniature Campaigns' in Comparison: The German Televised Debates, 2002-09, in: German Politics, 20/1 (2011), 75–91, hier 76 f.

3 Siehe hierzu auch *ibid.*: Miniature Campaigns, a.a.O., 76 f.

4 *Bachl, M.*: Analyse rezeptionsbegleitend gemessener Kandidatenbewertungen in TV-Duellen. Erweiterung etablierter Verfahren und Vorschlag einer Mehrebenenmodellierung, Berlin, 2014, hier 23.

5 *Maier, J./Faas, T.*: Miniature Campaigns, a.a.O., 76.

6 *Faas, T./Maier, J.*: Wahlkämpfe im Miniaturformat: Fernsehdebatten und ihre Wirkung am Beispiel des TV-Duells 2013 zwischen Angela Merkel und Peer Steinbrück, in: Information. Wissenschaft und Praxis, 65/3 (2014), 163–168.

Die Forschung zu TV-Duellen bedient sich in der jüngsten Vergangenheit vor allem rezeptionsbegleitender Messungen (sog. *Real-Time Response* oder RTR-Verfahren), bei denen Zuschauer ihren unmittelbaren Eindruck der Debatte wiedergeben können. Diese Verfahren sind wegen ihrer hohen Präzision besonders hilfreich, zugleich sind sie bislang aber nur in begrenztem Rahmen einsetzbar, da sie spezialisierte Messinstrumente voraussetzen (für gewöhnlich mit einem Drehknopf ausgestattete *Dialer*) und Untersuchungsteilnehmer deshalb zu einem labor-experimentellen Setting eingeladen werden müssen. Hier setzt das in diesem Beitrag vorgestellte Debat-O-Meter an: Als Gemeinschaftsprojekt von Politikwissenschaft und Informatik baut es auf einer an der Universität Freiburg entwickelten browserbasierten Software auf und erfasst die unmittelbaren Bewertungen der Zuschauer in Echtzeit. Das Debat-O-Meter (www.debatometer.de) ist als WebApp von jedem Smartphone (aber auch von Tablet-PCs, Notebooks oder Desktop-Rechnern) über das Internet erreichbar und ermöglicht eine sekundengenaue Datenerhebung, die verspricht, herkömmlichen RTR-Ansätzen vergleichbar zu sein. Forscher müssen mit dem Debat-O-Meter keine eigene Hardware mehr bereitstellen, sodass Untersuchungen nicht nur günstiger werden, sondern auch weitaus größere Stichproben verarbeiten können. Da Internetanschlüsse praktisch überall verfügbar sind, ist es im Hinblick auf die räumliche Repräsentation herkömmlichen Herangehensweisen sogar deutlich überlegen. Zugleich ermöglicht es Teilnehmern, eine Debatte von zu Hause aus zu bewerten, was ein deutlich lebensnäheres Umfeld darstellt als herkömmliche Laborsettings.

Im Zuge der Landtagswahl 2016 in Baden-Württemberg kam das Debat-O-Meter insgesamt drei Mal zum Einsatz: Erstmals beim TV-Duell am 14. Januar 2016 zwischen Ministerpräsident *Winfried Kretschmann* (Grüne) und seinem Herausforderer *Guido Wolf* (CDU). Nach Aussage des SWR verfolgten 460.000 Zuschauer die Diskussion, was einer Quote von 12,2% entspricht. Seinen zweiten Einsatz (hier aus Platzgründen⁷ nicht berichtet) fand es bei einer Podiumsdiskussion der Kandidaten von Grünen, CDU, SPD und FDP für den Wahlkreis Freiburg II am 28. Januar 2016. Dritter Anwendungstermin war der 10. März 2016 im Rahmen der *Elefantenrunde* im SWR, bei der die Spitzenkandidaten der sechs aussichtsreichsten Parteien debattierten und die von 550.000 Zuschauern (Quote: 13,6%) gesehen wurde.

7 Eine Kurzanalyse findet sich auf www.debatometer.de/analysen/Analyse_28_01_2016.pdf.

Der nachfolgende Beitrag verfolgt insgesamt drei Fragestellungen:

- (1) Die erste – methodologische – Frage zielt auf die Validität der RTR-Messung ab: Inwiefern ist es möglich, die in der Debattenforschung interessierenden Konstrukte mit dem Debat-O-Meter zu messen und welche Unterschiede und Gemeinsamkeiten ergeben sich zu den bisher eingesetzten physischen Geräten?
- (2) Die zweite Frage stellt auf die Debattenwahrnehmung ab: Wie werden die Inhalte der TV-Debatte von den Zuschauern wahrgenommen? Welche Wahrnehmungsmuster lassen sich allgemein erkennen und wie wirken die Kandidatenstrategien auf die unmittelbaren Bewertungen der Politiker?
- (3) Die dritte Fragestellung fokussiert auf die Debattenwirkung: Welche unmittelbaren Wirkungen hat das TV-Duell auf die Bewertung der Kandidaten sowie auf andere politische Einstellungen, wie z. B. die Wahlabsticht?

Der Beitrag beginnt mit einem Überblick über den Forschungsstand. Anschließend wird das Debat-O-Meter vorgestellt. Die folgenden Abschnitte widmen sich dann den drei Forschungsfragen. Schließlich werden die zentralen Befunde zusammengefasst und ein Ausblick auf zukünftige Anwendungen gegeben.

II. Forschungsstand

Die Politikwissenschaft hat sich mit TV-Duellen bereits seit deren Aufkommen Mitte der 1950er Jahre intensiv befasst.⁸ In ihrem 2004 erschienenen Überblick zählen *McKinney* und *Carlin* bereits über 800 Studien, der Großteil davon im Kontext der US-Präsidentenwahlen. Für die deutsche Forschung unterscheiden *Maier* und Kollegen drei große thematische Felder⁹: Erstens stellen viele Arbeiten auf den Inhalt des Duells ab und fragen, mit welchen Strategien Kandidaten zu punkten versuchen. Andere richten den Fokus auf die Berichterstattung in den Medien. Das dritte Feld untersucht dagegen motivationale, kognitive, affektive und verhaltensbezogene Effekte, die sich infolge des Duells einstellen.

8 Für einen Überblick über die US-Forschung siehe *McKinney, M./Warner, B.*: Do presidential debates matter?, in: *Argumentation and Advocacy*, 49/4 (2013), 238–258; *McKinney, M./Carlin, D.*: Political Campaign Debates, in: *Kaid, L. (Hg.): Handbook of Political Communication Research*, Mahwah, 2004, 203–234 oder *Benoit, W./Hansen, G./Verser, R.*: A Meta-Analysis of the Effects Viewing U.S. Presidential Debates, in: *Communication Monographs*, 70/4 (2003), 335–350. Für einen Überblick über die deutsche Forschung siehe *Maier, J./Faas, T./Maier, M.*, a.a.O.

9 *Maier, J./Faas, T./Maier, M.*, a.a.O., 41.

Für die USA hat die Forschung mehrere Wirkmechanismen isoliert. Dabei zeigt sich, dass TV-Debatten einerseits im Kontext des Wahlkampfes verstanden werden müssen und dass ihre Effekte andererseits davon abhängen, welche Voraussetzungen Zuschauer mitbringen. So haben Debatten vor allem dann einen Effekt, wenn ein Kandidat nur wenig bekannt ist, viele Wähler unentschieden sind, wenn der Abstand zwischen den Kandidaten gering ist oder wenn Parteibindungen schwach sind.¹⁰ Parteianhänger neigen dazu, sich in ihrer Entscheidung bestätigt zu sehen, während Meinungsänderungen bei ungebundenen aber interessierten Zuschauern am größten sind. Entsprechend wird TV-Duellen ein Einfluss auf die Wahl attestiert, zugleich aber auch dessen Grenzen betont, vor allem bei der Wahlentscheidung selbst¹¹, da politische Prädispositionen wie die Parteiidentifikation stark beeinflussen, wie die Debatte wahrgenommen und bewertet wird.¹² Gleichwohl lässt sich zeigen, dass TV-Duelle die Wahrnehmung eines Kandidaten durchaus beeinflussen und dass sich Entscheidungen Ungebundener anhand von ihnen konkretisieren und auch ändern.¹³ Allerdings ist die Stabilität vieler Effekte insofern unsicher, als Duelle im Rahmen eines Wahlkampfes ablaufen, der seinerseits auf das Duell reagiert.¹⁴ Auch dienen TV-Duelle Zuschauern dazu, sich Wissen über die Kandidaten anzueignen, allerdings ist der Effekt für unbekanntere Kandidaten größer und es kommt zu Sättigungseffekten.¹⁵ Zudem profitieren meist beide Teilnehmer, ohne dass der Sieg des Einen notwendigerweise auf Kosten des Anderen erfolgt.¹⁶ Darüber hinaus steigern TV-Duelle u. a. das subjektive Kompetenzgefühl, das Interesse am Wahlkampf sowie die Wahlbeteiligung.¹⁷

- 10 *McKinney, M.*: Debates, in: Kaid, L./Holtz-Bacha, C. (Hg.): *Encyclopedia of political communication*, Thousand Oaks, 2008, 159–165, hier 163.
- 11 Siehe z. B. *McKinney, M./Warner, B.*, a.a.O., 239 f. und *Blais, A./Boyer, M.*: Assessing the Impact of Televised Debates: The Case of the 1988 Canadian Election, in: *British Journal of Political Science*, 26/2 (1996), 143–164.
- 12 *Schrott, P./Lanoue, D.*: The power and limitations of televised presidential debates: Assessing the real impact of candidate performance on public opinion and vote choice, in: *Electoral Studies*, 32/4 (2013), 684–692.
- 13 Siehe z. B. *Pfau, M./Cho, J./Chong, K.*: Communication Forms in U.S. Presidential Campaigns: Influences on Candidate Perceptions and the Democratic Process, in: *The Harvard International Journal of Press/Politics*, 6/4 (2001), 88–105; *Yawn, M./Ellsworth, K./Beatty, B./Kahn, K.*: How a Presidential Primary Debate Changed Attitudes of Audience Members, in: *Political Behavior*, 20/2, (1998), 155–181.
- 14 *McKinney, M./Carlin, D.*, a.a.O.
- 15 *Holbrook, T.*: Political Learning from Presidential Debates, in: *Political Behavior*, 21/1 (1999), 67–89; *Benoit, W./McKinney, M./Stephenson, M.*: Effects of Watching Primary Debates in the 2000 U.S. Presidential Campaign, in: *Journal of Communication*, 52/2 (2002), 316–331; *Benoit, W./Hansen, G.*: Presidential Debate Watching, Issue Knowledge, Character Evaluation, and Vote Choice, in: *Human Communication Research*, 30/1 (2004), 121–144.
- 16 *McKinney, M./Warner, B.*, a.a.O.; *Blais, A./Perrella, A.*: Systemic Effects of Televised Candidates' Debates, in: *The International Journal of Press/Politics*, 13/4 (2008), 451–464.
- 17 *McKinney, M./Rill, L./Gully, D.*: Civic engagement through presidential debates: Young citizens attitudes of political engagement throughout the 2008 election, in: *McKinney, M./Banwart, M.* (Hg.): *Communication in the 2008 U.S. election: Digital natives elect a president*, New York, 2011, 121–141; *Wald, K./*

Auch für Deutschland analysiert eine Vielzahl innovativer Arbeiten TV-Duelle.¹⁸ Im Großen und Ganzen konnten diese Arbeiten zeigen, dass viele der für die USA beobachteten Zusammenhänge hierzulande in ähnlicher Weise auftreten. So ziehen TV-Duelle vor allem politisch Involvierte an, erreichen sie ein deutlich größeres Publikum als andere Formate und damit auch vielfach politisch weniger Interessierte.¹⁹ Auch für Deutschland spielt die Parteiidentifikation eine zentrale Rolle und mit Blick auf die Wahlentscheidung nutzen Zuschauer ein TV-Duell auf vielfältige Weise: Während Unentschiedene Präferenzen ausbilden, bestätigen Entschiedene eher ihre Wahl, in geringerer Anzahl finden sich aber auch Präferenzänderungen.²⁰

Strukturell beeinflussen TV-Duelle die Wahl indirekt über eine Veränderung der Kandidatenorientierung, die sich einerseits aus zuvor bestehenden Präferenzen sowie andererseits aus der Debattenleistung speist (dabei ist die Wahrnehmung, eine Debatte gewonnen zu haben, weitgehend eine Funktion der laufenden Bewertung

Lupfer, M.: The Presidential Debate As a Civics Lesson, in: *Public Opinion Quarterly*, 42/3 (1978), 342–353; *McKinney, M./Warner, B.*, a.a.O.; *McKinney, M./Chattopadhyay, S.*: Political Engagement through Debates: Young Citizens' Reactions to the 2004 Presidential Debates, in: *American Behavioral Scientist*, 50/9 (2007), 1169–1182.

- 18 *Faas, T./Maier, J.*: Mobilisierung, Verstärkung, Konversion? Ergebnisse eines Experiments zur Wahrnehmung der Fernsehduelle im Vorfeld der Bundestagswahl 2002, in: *Politische Vierteljahresschrift*, 45/1 (2004), 55–72; *Maier, J./Faas, T.*: Schröder gegen Stoiber. Wahrnehmung, Verarbeitung und Wirkung der Fernsehdebatten im Bundestagswahlkampf 2002, in: *Falter, J./Gabriel, O./Weßels, B.* (Hg.): *Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 2002*, Wiesbaden, 2005, 77–101; *Maier, J./Faas, T./Maier, M.*, a.a.O.; *Maurer, M./Reinemann, C.*: Schröder gegen Stoiber: Nutzung, Wahrnehmung und Wirkung der TV-Duelle, Wiesbaden, 2003; *Maurer, M./Reinemann, C./Maier, J./Maier, M.* (Hg.): *Schröder gegen Merkel. Wahrnehmung und Wirkung des TV-Duells 2005 im Ost-West-Vergleich*, Wiesbaden, 2007; *Bachl, M./Brettschneider, F./Ottler, S.* (Hg.): *Das TV-Duell in Baden-Württemberg 2011. Inhalte, Wahrnehmungen und Wirkungen*, Wiesbaden, 2013; *Maier, J./Faas T.*: *Miniature Campaigns*, a.a.O.; *Klein, M.*: Die TV-Duelle: Events ohne Effekt?, in: *Güllner, M./Dülmer, H./Klein, M./Ohr, D./Quandt, M./Rosar, U./Klingemann, H.-D.* (Hg.): *Die Bundestagswahl 2002. Eine Untersuchung im Zeichen hoher politischer Dynamik*, Wiesbaden, 2005, 143–159; *Klein, M./Rosar, U.*: Wirkungen des TV-Duells im Vorfeld der Bundestagswahl 2005 auf die Wahlentscheidung, in: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 59/1 (2007), 81–104.
- 19 *Faas, T./Maier, J.*: Chancellor-Candidates in the 2002 Televised Debates, in: *German Politics*, 13/2 (2004), 300–316; *Maier, J./Faas, T.*: *Miniature Campaigns*, a.a.O.; *Maier, J./Faas, T.*: *Schröder gegen Stoiber*, a.a.O.
- 20 *Maier, J./Faas, T.*: *Miniature Campaigns*, a.a.O.; *Klein, M.*, a.a.O.; *Klein, M./Pötschke, M.*: Haben die beiden TV-Duelle im Vorfeld der Bundestagswahl 2002 den Wahlausgang beeinflusst? Eine Mehrebenenanalyse auf der Grundlage eines 11-Wellen-Kurzfristpanels, in: *Falter, J./Gabriel, O./Weßels, B.* (Hg.): *Wahlen und Wähler. Analysen aus Anlass der Bundestagswahl 2002*, Wiesbaden, 2005, 357–370; *Faas, T./Maier, J.*: *Chancellor-Candidates*, a.a.O.; *Maier, J.*: Erfolgreiche Überzeugungsarbeit. Urteile über den Debattensieger und die Veränderung der Kanzlerpräferenz, in: *Maurer, M./Reinemann, C./Maier, J./Maier, M.* (Hg.): *Schröder gegen Merkel. Wahrnehmung und Wirkung des TV-Duells 2005 im Ost-West-Vergleich*, Wiesbaden, 2007, 91–109; *Maier, M.*: Verstärkung, Mobilisierung, Konversion. Wirkungen des TV-Duells auf die Wahlabsicht, in: *Maurer, M./Reinemann, C./Maier, J./Maier, M.* (Hg.): *Schröder gegen Merkel. Wahrnehmung und Wirkung des TV-Duells 2005 im Ost-West-Vergleich*, Wiesbaden, 2007, 145–165.

während der Diskussion, auch wenn Parteianhänger meist einen Sieg ihres Kandidaten erwarten).²¹ Gleichwohl ist die Wahrnehmung der Diskussion selbst aber deutlich durch die Parteiidentifikation vorgeformt, obwohl Anhänger trotz aller Skepsis natürlich auch Sympathien für die andere Seite entwickeln können.²² Umgekehrt bedeutet dies, dass die Effekte von TV-Duellen bei Ungebundenen am stärksten sind.²³ Kandidaten können vor allem mit Aussagen zu Themen punkten, für die ihre Partei als kompetent gilt, auch finden eher allgemeine Aussagen ebenso wie populäre bis populistische Statements breite Zustimmung.²⁴ Konkrete Politik- oder Personalausagen wirken dagegen eher polarisierend, ähnlich wie Angriffe auf den politischen Gegner, wie interessanterweise auch die damit verbundene Defensive abgelehnt wird; Selbstpräsentationen dagegen kommen insbesondere bei Ungebundenen gut an.²⁵ Entgegen populärer Erwartung spielen non-verbale Aussagen eine eher kleine Rolle.²⁶

III. Das Debat-O-Meter

Seit einigen Jahren werden bei der Erforschung von TV-Duellen rezeptionsbegleitende RTR-Messverfahren²⁷ eingesetzt. Bei diesen geben Zuschauer kontinuierlich Auskunft über ein meist globales Konstrukt, wie z. B. ihre Einschätzung der

- 21 Z. B. Maier, J., a.a.O.; Maier, J./Faas, T.: Debattenwahrnehmung und Kandidatenorientierung. Eine Analyse von Real-Time-Response- und Paneldaten zu den Fernsehduellen im Bundestagswahlkampf 2002, in: Zeitschrift für Medienpsychologie, 16/1 (2004), 26–35. Die Wahrnehmung als Sieger entfaltet auch Tage nach dem Duell noch eine Wirkung, siehe z. B. Klein, M., a.a.O.; Klein, M./Pötschke, M., a.a.O.; Maier, J./Faas, T.: Schröder gegen Stoiber, a.a.O.; Faas, T./Maier, J.: Chancellor-Candidates, a.a.O.
- 22 Maier, J./Faas, T./Maier, M., a.a.O.; Faas, T./Maier, J.: Mobilisierung, Verstärkung, Konversion, a.a.O.
- 23 Maier, J./Faas, T.: Schröder gegen Stoiber, a.a.O.; Maier, J./Faas, T.: Miniature Campaigns, a.a.O.; Faas, T./Maier, J.: Chancellor-Candidates, a.a.O.
- 24 Zum Vorteil bei Issue Ownership siehe z. B. Bachl, M.: Die Wahrnehmung des TV-Duells, in: Bachl, M./Brettschneider, F./Ottler, S. (Hg.): Das TV-Duell in Baden-Württemberg 2011. Inhalte, Wahrnehmungen und Wirkungen, Wiesbaden, 2013, 135–169. Zu Statements siehe Reinemann, C./Maurer, M.: Populistisch und unkonkret. Die unmittelbare Wahrnehmung des TV-Duells, in: Maurer, M./Reinemann, C./Maier, J./Maier, M. (Hg.): Schröder gegen Merkel. Wahrnehmung und Wirkung des TV-Duells 2005 im Ost-West-Vergleich, Wiesbaden, 2007, 53–89.
- 25 Reinemann, C./Maurer, M.: Unifying or Polarizing? Short-Term Effects and Postdebate Consequences of Different Rhetorical Strategies in Televised Debates, in: Journal of Communication, 55/4 (2005), 775–794; aber Bachl, M.: Wahrnehmung des TV-Duells, a.a.O.; Bachl, M.: Kandidatenbewertungen in TV-Duellen, a.a.O.; Reinemann, M./Maurer, C.: Populistisch und unkonkret, a.a.O.
- 26 Nagel, F.: Die Wirkung verbaler und nonverbaler Kommunikation in TV-Duellen. Eine Untersuchung am Beispiel von Gerhard Schröder und Angela Merkel, Wiesbaden, 2012.
- 27 Für eine Einführung siehe Maurer, M.: Real-Time Response Messung: Kontinuierliche Befragung in Echtzeit, in: Möhring, W./Schlütz, D. (Hg.): Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft, Wiesbaden, 2013, 219–234 oder Biocca, F./David, P./West, M.: Continuous response measurement (CRM): A computerized tool for research on the cognitive processing of communication messages, in: Lang, A. (Hg.): Measuring psychological responses to media messages, Hillsdale, 1994, 15–64. Für einen Überblick über TV-Duell-Studien mit RTR-Messungen siehe Bachl, M.: Kandidatenbewertungen in TV-Duellen, a.a.O.

Kandidaten. Üblicherweise werden zur Messung physische Geräte benutzt, an denen Teilnehmer mithilfe von Knöpfen, Schub- oder Drehreglern ihr Urteil sekundengenau abgeben können. RTR-Studien fokussieren meist auf die Frage, welchen Effekt Aspekte der Debatte auf das gemessene Konstrukt haben, beispielsweise welche Aussagetypen beim Publikum erfolgreich sind oder wie ein Zuschauer zur Wahrnehmung eines Debattensiegers kommt. RTR-Messungen haben die Forschung zu TV-Duellen stark bereichert, erlauben sie doch einen sehr detaillierten Einblick, wie ein Publikum eine Diskussion wahrnimmt. Ein großer Nachteil der Methode ist jedoch noch, dass für die Erhebung physische Geräte benötigt werden, was wiederum bedeutet, dass Teilnehmer nur in einem Laborsetting untersucht werden können. Direkte Folge daraus ist, dass weitgehend offen bleibt, ob und wie gut sich Befunde auf alltägliche Situationen vor dem heimischen Fernseher übertragen lassen. Vier Felder sind in dieser Hinsicht evident:²⁸

- (1) Die Anforderungen an die Teilnehmer (hohe Motivation, Anfahrt usw.) begünstigen politisch stark Interessierte, die sich aber mit Blick auf politische Informiertheit, die Klarheit und Stabilität von Orientierungen sowie in der Art der politischen Informationsverarbeitung vom Rest der Bevölkerung unterscheiden.²⁹
- (2) Politische Involviertheit und die damit verbundene Sicherheit im Umgang mit politischer Bewertung moderiert die Wirkstruktur von TV-Duellen. Im Sinne der *social facilitation theory*³⁰ – die grob besagt, dass einfache Aufgaben unter Anwesenheit anderer leichter, komplexe dagegen schwerer fallen – könnte die *Laborüberwachung* bewirken, dass die Konsistenz der Urteile Uninvolverter gegenüber dem privaten Umfeld nachlässt, während sie für stark Involvierte steigt.
- (3) Teilnehmer von RTR-Messungen reagieren im Labor mitunter aufeinander und passen sich aneinander bzw. an ein globales Meinungsklima an.³¹

28 *Ibid.*: Kandidatenbewertungen in TV-Duellen, a.a.O., hier 53–58.

29 Siehe hierzu *Converse, P.*: Assessing the capacity of mass electorates, in: *Annual Review of Political Science*, 3 (2000), 331–353; *ibid.*: The nature of belief systems in mass publics (1964), in: *Critical Review*, 18/1 (2006), 1–74; *Campbell, A./Converse, P./Miller, W./Stokes, D.*: *The American voter*, New York, 1960, hier Kap. 10; *Zaller, J.*: *The Nature and Origins of Mass Opinion*, Cambridge, 1992.

30 *Zajonc, R.*: Social facilitation, in: *Science*, 149/3681 (1965), 269–274; *Markus, H.*: The Effect of Mere Presence on Social Facilitation: An Unobtrusive Test, in: *Journal of Experimental Social Psychology*, 14/4 (1978), 389–297; *Aiello, J./Douthitt, E.*: Social Facilitation from Triplett to Electronic Performance Monitoring, in: *Group Dynamics: Theory, Research, and Practice*, 5/3 (2001), 163–180. Siehe auch *Papastefanou, G.*: Reliability and Validity of RTR Measurement Devices, GESIS Working Papers 27, Mannheim, 2013.

31 *Ramanathan, S./McGill, A.*: Consuming with Others: Social Influences on Moment-to-Moment and Retrospective Evaluations of an Experience, in: *Journal of Consumer Research*, 34/4 (2007), 506–524; *Wolf,*

Da Teilnehmer oft per Quotenplan ausgewählt werden, während private Mit-Zuschauer meist aus einem sozial und politisch homogenen Umfeld stammen, könnten sich die Effekte zwischen Labor und realem Setting unterscheiden, zumal sich bereits Hinweise auf eine soziale Einbettung der Wirkung von Duellen finden.³²

- (4) Bisherige RTR-Studien unterliegen oft praktischen Begrenzungen, da meist nur 100-120 Personen teilnehmen können. Zwar versuchen bestehende Studien dieses Manko über Quotenpläne zu korrigieren, doch kann das nur mit einer begrenzten Anzahl von Kriterien erfolgen und stellt sich das Problem, die Felder des Quotenplans auch wirklich alle zu besetzen.³³

Alle vier Punkte beeinflussen bisher die externe Validität von RTR-Studien negativ. Das Debat-O-Meter ist ein Versuch, diese Lücke zu schließen. Als WebApp für Smartphones und Tablets kann es ohne spezialisierte Hardware eingesetzt werden. Neben größeren Teilnehmerzahlen beabsichtigt es vor allem, das bisher einschränkende Laborsetting zu verlassen. Der gegenwärtige Prototyp (siehe *Abb. 1*) entstand aus dem *Classroom Response System* SMILE³⁴, das am Freiburger Lehrstuhl für Rechnerarchitektur zur Unterstützung der Lehre entwickelt wurde.

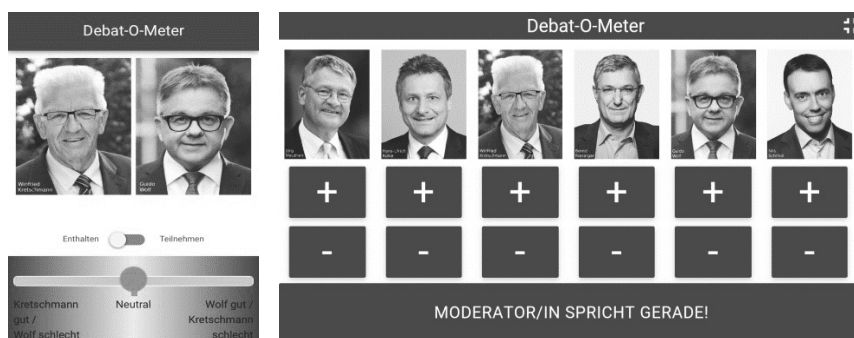
Auf Seiten des Benutzers (*Client*) ist das Debat-O-Meter als Website implementiert. Das bedeutet, dass Nutzer eine Internetseite mit dem Browser öffnen, deren Design und Benutzung einer *herkömmlichen* Smartphone-App entspricht. Für die Software-Entwicklung hat dies den ganz praktischen Vorteil, dass man zwar nicht ganz die gleichen Freiheiten wie bei der Programmierung einer normalen App besitzt, dafür aber die Anwendung nur mit einer überschaubaren Anzahl von Browsern testen muss, was wiederum die Entwicklung beschleunigt. Bei einer herkömmlichen App muss dagegen die Nutzbarkeit auf zahllosen Geräten geprüft

B.: Beurteilung politischer Kandidaten in TV-Duellen: Effekte rezeptionsbegleitender Fremdmeinungen auf Zuschauerurteile, Baden-Baden, 2010; *Fein, S./Goethals, G./Kugler, M.*: Social Influence on Political Judgments: The Case of Presidential Debates. How do citizens respond to campaign events?, in: *Political Psychology*, 28/2 (2007), 165–192.

- 32 *McPherson, M./Smith-Lovin, L./Cook, J.*: Birds of a Feather: Homophily in Social Networks, in: *Annual Review of Sociology*, 27 (2001), 415–444; *Mutz, D.*: Hearing the Other Side: Deliberative versus Participatory Democracy, Cambridge, 2006; *Maier, J./Faas, T.*: The affected German voter: Televised debates, follow-up communication and candidate evaluations, in: *Communications*, 28/4 (2003), 383–404.
- 33 *Bachl, M./Brettschneider, F./Ottler, S.*: Die TV-Duell-Studie Baden-Württemberg 2011, in: dies. (Hg.): *Das TV-Duell in Baden-Württemberg 2011*, Wiesbaden, 2013, 7–27.
- 34 *Feiten, L./Bührer, M./Sester, S./Becker, B.*: SMILE – SMARTPHONES IN LECTURES. Initiating a Smartphone-based Audience Response System as a Student Project, in: Helfert, M./Martins, M.J./Cordeiro, J. (Hg.): *Proceedings of the 4th International Conference on Computer Supported Education*, 2012, 288–293; *Feiten, L./Weber, K./Becker, B.*: SMILE: Smartphones in der Lehre – ein Rück- und Überblick 2013, in: Horbach, M. (Hg.): *INFORMATIK 2013: Informatik angepasst an Mensch, Organisation und Umwelt*, Bonn, 2013, 255–269.

werden, was oft die Beauftragung spezialisierter Testfirmen notwendig macht, welche die entsprechenden Geräte vorhalten können. Zudem zeigte sich bei der Konzeption des Debat-O-Meters, dass das benötigte Funktionsrepertoire auch problemlos mit den Techniken klassischer Webprogrammierung (HTML5, Javascript) realisiert werden kann. Zusätzlich hat die gewählte Lösung den (wahrscheinlich noch bedeutenderen) Vorteil, dass Nutzer die App nicht auf ihrem Smartphone installieren, sondern lediglich eine Website besuchen müssen, was ein potenzielles Hindernis für eine Teilnahme umgeht.

Abbildung 1: Oberflächen des Debat-O-Meters



Oberfläche des Debat-O-Meters am 14.01.2016 und 10.03.2016. *Quelle:* Eigene Darstellung.

Auf Seiten des Servers folgt die Struktur des Debat-O-Meters einer klassischen LAMP-Infrastruktur, d. h. der Server wird mit Linux betrieben, setzt die Software Apache zur Internetkommunikation ein und speichert die gesammelten Daten in einer MySQL-Datenbank, die über die Skriptsprache PHP bedient wird. Der Debat-O-Meter-Server wird nur dann aktiv, wenn ein Nutzer über seinen Client eine Nachricht (z. B. eine Bewertung durch Bewegen des Sliders oder Drücken eines Buttons) verschickt. Die Nachricht enthält dabei den Wert der Sliderposition oder des Buttons. Der Server versieht diesen Wert beim Eintreffen mit der aktuellen Uhrzeit und speichert beides zusammen mit der Kennung des Nutzers in der Datenbank. Aus diesen Rohdaten werden dann alle Auswertungen erstellt.

Die gegenwärtige Version des Debat-O-Meters berücksichtigt noch nicht den Zeitraum, der zwischen Absenden und Eintreffen einer Nachricht vergehen kann. Daher wurden die unten geschilderten Versuche noch in einem Raum (so entfallen Unterschiede im Timing des TV-Signals) und mit hochperformantem WLAN durchgeführt, sodass eventuelle Verzögerungen, die eine Synchronisation der RTR-Signale notwendig machen würden, vernachlässigt werden können. Für spätere Versionen, die dann von überall aus im Internet genutzt werden sollen, muss

dagegen noch sichergestellt werden, dass eine Bewertung der korrekten Stelle in der Diskussion zugeordnet werden kann. Derzeit werden hierzu heuristische Ansätze (meist erfolgt z. B. eine Bewertung der sprechenden Person), implizite manuelle Codierung durch die Nutzer selbst (z. B. Drücken eines Buttons zu wohl definierten Zeitpunkten) sowie autonome Ansätze (z. B. Reaktion des Clients auf Audio-Signale aus der Debatte) untersucht. Bei letzteren ist insbesondere darauf zu achten, dass bei der Umsetzung keine Information übertragen wird, die in die Privatsphäre des Nutzers eingreift.

Im Folgenden stellen wir den bisherigen Stand der Arbeiten zur Etablierung des Debat-O-Meters als smartphonebasiertes RTR-Instrument dar. Wir tun dies anhand zweier Laborstudien, die zur Landtagswahl Baden-Württemberg 2016 an der Universität Freiburg durchgeführt wurden. Da das Publikum jeweils aus Studierenden und Interessierten frei rekrutiert wurde, sind die Ergebnisse nicht repräsentativ und noch nicht verallgemeinerbar, sie stellen jedoch den *Proof of Concept* dar, nach dem TV-Debatten mit einem virtuellen Instrument vermessen werden können.

IV. Die Debattenwahrnehmung

Im Vorfeld der Landtagswahl fanden zwei große TV-Diskussionen statt, die mit dem Debat-O-Meter begleitet wurden. Beide Diskussionen kamen erst nach längerer Auseinandersetzung zustande, was die komplexe Situation im Vorfeld der Wahl illustriert: Nachdem CDU-Spitzenkandidat *Guido Wolf* ein *Triell* mit den Koalitionären *Winfried Kretschmann* (Grüne) und *Nils Schmid* (SPD) unter Verweis auf ein Fehlen der FDP abgelehnt hatte, kam es am 14. Januar 2016 zu einem Aufeinandertreffen zwischen *Kretschmann* und *Wolf* als Streitgespräch in der Sendung *Zur Sache! Baden-Württemberg*. Kurz darauf, am 20. Januar 2016, fand im Theaterhaus Stuttgart ein zweites Duell statt, welches allerdings nur live im Internet übertragen wurde. Eine Zweitausstrahlung fand jedoch über regionale Fernsehstationen einige Tage später statt. Die *Elefantenrunde* als zweite große Diskussion am 10. März 2016 hätte nach Wunsch des SWR mit AfD und Linke stattfinden sollen, jedoch lehnten *Kretschmann* und *Schmid* eine Beteiligung erst ab, um der AfD keine Bühne zu bieten, willigten dann aber nach großem öffentlichen Druck in eine Diskussion ein. Im Folgenden werden die Debat-O-Meter-Befunde zu beiden Debatten in Kurzform dargestellt, für das Duell wird zudem ein Vergleich zwischen dem Debat-O-Meter und einer parallelen Messung mit Dialern gezogen.

1. Das TV-Duell Winfried Kretschmann/Guido Wolf am 14. Januar 2016

Der erste Versuch fand in einem Hörsaal der Universität Freiburg statt. Insgesamt erschienen rund 250 Personen, um die Diskussion live zu sehen. Die Rekrutierung erfolgte ohne Quotenplan über Ankündigungen in Vorlesungen, über Flyer, Plakate, Online-Werbung sowie über Medienberichte, sodass das Publikum im weiteren Sinn das Umfeld der Universität widerspiegelte und die Ergebnisse nicht als allgemeines Stimmungsbild interpretiert werden sollten. Die Teilnahme am Debat-O-Meter war optional, gleichwohl beteiligte sich ein großer Teil der Besucher. Bei Eintreffen am Hörsaal wurden 35% der interessierten Teilnehmer randomisiert in eine Dialer-Gruppe³⁵ eingeteilt, der Rest wurde der Debat-O-Meter-Gruppe zugeschlagen (Personen ohne Smartphone erhielten ein Leihgerät). Beide Gruppen wurden zu einer identischen Vor- und Nachbefragung gebeten und sollten für die RTR-Bewertung ihren allgemeinen Eindruck der beiden Kandidaten saldieren, entweder über Dialer oder über einen Slider auf dem Debat-O-Meter (vgl. *Abb. 1*). Da sich nicht alle Personen nach der Befragung auch am Debat-O-Meter beteiligten und einzelne die Veranstaltung früher verließen, unterscheiden sich die Fallzahlen zwischen den Stufen des Versuchs. Die Dialer-Gruppe umfasste insgesamt 50 Personen (darin enthalten sind 12 Personen, die aufgrund einer kurzzeitigen Überlastung des Webservers aus der Debat-O-Meter-Gruppe in die Dialer-Gruppe wechselten, um die Infrastruktur zu entlasten). Im Debat-O-Meter sind 110 RTR-Kurven erfasst. Für die Vor- und Nachbefragung liegen 200 (149 für die Debat-O-Meter-Gruppe bzw. 41 für die Dialer-Gruppe) respektive 194 (151 zu 43) Fragebögen vor, für 184 (150 zu 34) Personen liegen beide Fragebögen vor. Nach Sichtprüfung wurden 6 bzw. 18 Kurven aussortiert, sodass die RTR-Kurven für die Dialer auf 44 Fällen beruhen, während im Debat-O-Meter 92 Fälle in die Analyse eingingen.³⁶

Entsprechend der offenen Rekrutierung bestand das Publikum zu 62% aus Studierenden und Schülern, aber auch Erwerbstätige (27%) und Rentner (5%) waren anwesend (Anteile jeweils an den gültigen Antworten). Rund zwei Drittel waren Männer, 41% gaben in der Vorbefragung an, die Grünen wählen zu wollen, je rund 8% nannten CDU und SPD, 7% die Linke, knapp 25% waren noch unentschieden. Rund 68% würden bei einer Direktwahl des Regierungschefs *Winfried*

35 Das konkret verwendete Gerät war ein *Perception Analyzer* (Model V) des Herstellers Dialsmith (www.dialsmith.com).

36 Hierbei ist zu beachten, dass die Zuordnung der RTR-Kurven nicht über die Fragebögen, sondern über die jeweilige Auswertungssoftware erfolgte.

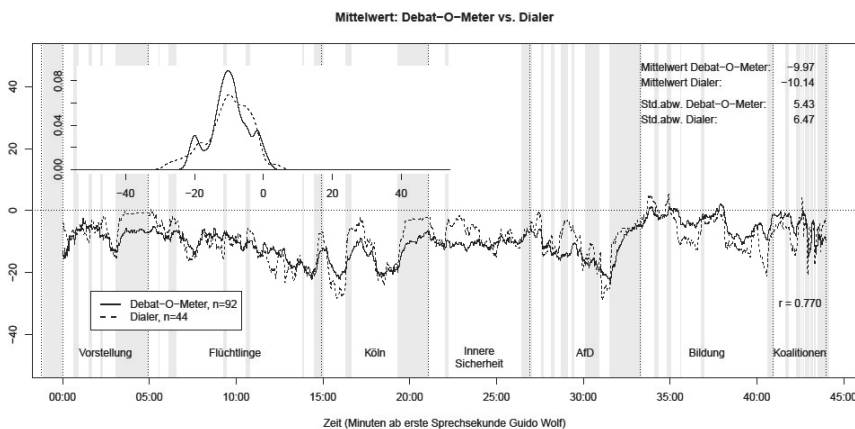
Kretschmann wählen, gut 6% *Guido Wolf*, der Rest gab an, keinen von beiden wählen zu wollen.

a) Die Messung mit dem Debat-O-Meter und mit Dialern

Unsere erste Frage stellt auf die Validität der RTR-Messung ab: Inwiefern ist es möglich, die in der Debattenforschung interessierenden Konstrukte auch mit dem Debat-O-Meter zu messen? Um diese Frage zu beantworten, wurden die Messwerte des Debat-O-Meters und der Dialer jeweils über alle Teilnehmer in der Gruppe gemittelt und verglichen. Durch die randomisierte Einteilung sollen sich die Kurven nicht wesentlich unterscheiden, sodass ein Vergleich auch ohne Einbeziehung der inhaltlichen Ebene und unabhängig von der konkreten Zusammensetzung des Publikums möglich ist. Das Ergebnis ist in *Abb. 2* dargestellt. Für alle berechneten Größen wurden die (in der Grafik grau hinterlegten) Sprechphasen der Moderation ausgeschlossen.

Wie *Abb. 2* zeigt, ähneln sich die Messreihen der beiden Methoden deutlich. Beide Kurven zeigen nur minimale Unterschiede in der Lage (-9,97 vs. -10,14), was darauf hindeutet, dass die Dialer und das Debat-O-Meter den gleichen Tenor der Wahrnehmung erfassen (die Skala reicht von -50 bei einem sehr guten/schlechten Eindruck von *Kretschmann/Wolf* bis +50 für die umgekehrte Konstellation). Offenbar haben beide Gruppen den Amtsinhaber vorne gesehen und dies auf beiden Messungen im gleichen Grad zum Ausdruck gebracht.

Abbildung 2: Debat-O-Meter versus Dialer



Quelle: Eigene Darstellung.

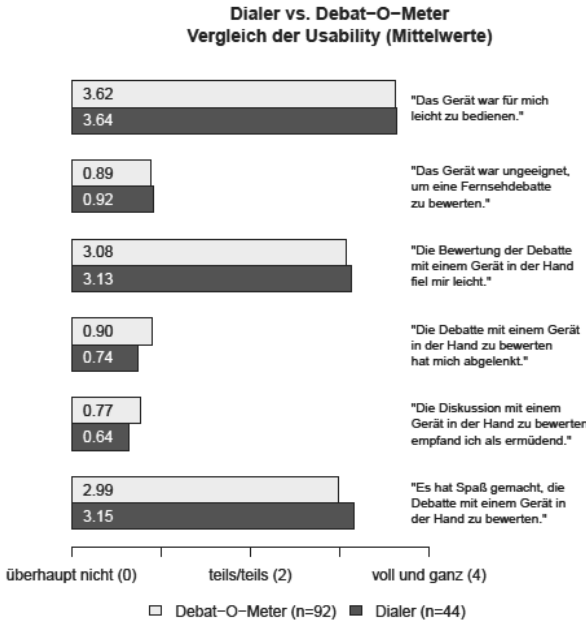
Im Hinblick auf die Standardabweichung zeigt sind, dass diese für die Dialer etwas erhöht ist (6,47 gegenüber 5,43), was vermutlich darauf zurückzuführen ist, dass sich Dialer ohne Hinsehen bedienen lassen, was wiederum eine höhere Bewertungsfrequenz möglich macht. Deutlich zu sehen ist die stärkere Volatilität der Dialer (auch wenn die konkrete Stelle nicht in die Berechnung mit einfließt), beispielsweise in der Moderationsphase vor Beginn der Diskussion zur Flüchtlingskrise (bis zum Beginn der Minute 5; die Sekunde 0:00 ist jene, in der *Guido Wolf* mit seiner ersten Antwort beginnt). Für alle Teilnehmer war die Messanweisung, den Regler in Moderationsphasen auf neutral zu stellen; die Dialer-Gruppe kam dieser Aufforderung jedoch (zumindest zu Beginn) eher nach als die Debat-O-Meter-Gruppe, allerdings relativieren sich die Unterschiede im späteren Diskussionsverlauf (siehe die Anmoderation zum Themenbereich Bildung).

Die etwas stärkere Reaktion der Dialer wird auch in der Kerndichteschätzung der Messwerte an den weiter ausgreifenden Rändern sichtbar (oberer Einsatz in *Abb. 2*). Andererseits wird aber auch deutlich, dass nicht nur die Gipfel beider Verteilungen an der gleichen Position liegen, sondern dass auch der zweite Gipfel, der einen Vorteil für *Kretschmann* anzeigt, noch erfasst wird. In der Detailsicht verlaufen beide Kurven deutlich parallel, oftmals sind in beiden Kurven die gleichen Veränderungen (mit unterschiedlicher Intensität) sichtbar. Entsprechend sind beide Kurven auch in dieser naiven Betrachtungsweise und ohne Glättung bereits stark korreliert ($r = 0,770$). Offenbar erfassen beide Messmethoden in weiten Teilen das gleiche interessierende Konstrukt.

Wie aber erscheint die Nutzung des Debat-O-Meters im Vergleich zu den Dialern aus Sicht der Teilnehmer? Eine Gegenüberstellung der Bedienbarkeit und Geeignetheit beider Instrumente findet sich in *Abb. 3*.

Gemittelt über die Antworten der beiden Gruppen zeigt sich, dass Debat-O-Meter und Dialer praktisch äquivalent abschneiden. Für die Anwender waren beide Geräte gleichermaßen einfach zu bedienen und schienen ihnen auch subjektiv geeignet, um damit eine Diskussion zu bewerten. Das eigene Urteil über die Diskussion zum Ausdruck zu bringen, war auf beiden Geräten gleichermaßen leicht, auch wenn die Bewertung auf dem Debat-O-Meter vielleicht geringfügig mehr Konzentration von den Teilnehmern erforderte als der Drehknopf. Im Fazit scheinen sich jedoch beide Gruppen einigermaßen einig zu sein, dass die Nutzung des Debat-O-Meters ähnlich angenehm ist wie die von Dialern. Entsprechend scheint es gangbar, RTR-Daten auch über Smartphones zu sammeln, ohne größere Einbußen im Hinblick auf die Usability hinnehmen zu müssen.

Abbildung 3: Usability Debat-O-Meter und Dialer



Quelle: Eigene Darstellung.

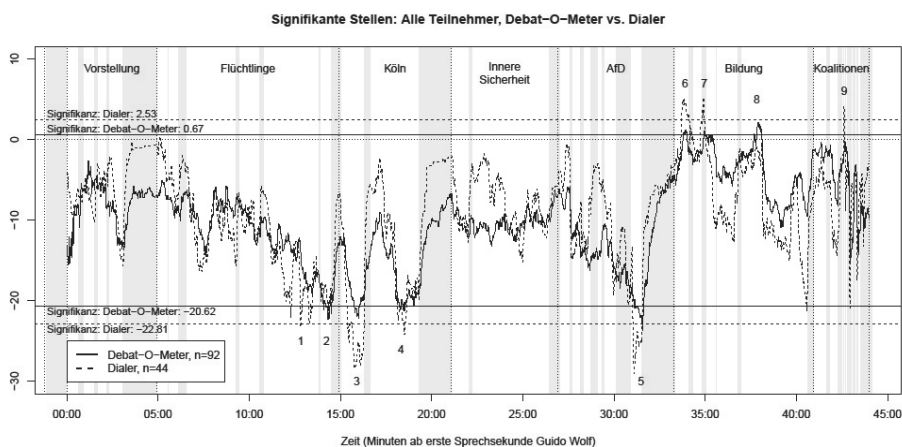
b) Die Debattenwahrnehmung

Ein dritter Aspekt des Vergleichs zwischen der etablierten Messung mit Dialern und dem Debat-O-Meter kreist um die Frage, ob die Analyse auf Grundlage der Daten aus beiden Quellen zum gleichen Ergebnis kommt, sprich: Sind beide Verfahren äquivalent, dann sollten sie sich nicht nur aus Sicht der produzierten Daten und aus Sicht der Anwender *nicht* unterscheiden, auch die aus der Messung gezogenen Schlüsse sollten im Grundsatz die gleichen sein. Diese Perspektive berührt nun erstmals auch die inhaltliche Ebene der Diskussion, sodass sie gleichermaßen für die erste und zweite Forschungsfrage relevant ist. *Abb. 4* zeigt den Verlauf der mittleren Bewertung noch einmal im Vergleich, diesmal aber aufgeschlüsselt danach, welche Stellen vom Publikum auffällig intensiv bewertet wurden. Um diese zu identifizieren ziehen wir das in der Literatur verbreitete Kriterium von *Biocca* und Kollegen³⁷ heran, jene Passagen als wichtig anzusehen, deren Bewertung mehr als 1,96 Standardabweichungen vom Mittelwert der Kurve entfernt sind. Für

37 *Biocca, F./David, P./West, M., a.a.O.* Für einen Überblick über solche Verfahren siehe auch *Bachl, M.: Kandidatenbewertungen in TV-Duellen, a.a.O., hier Kap. 5.2.*

das Debat-O-Meter liegen diese Grenzen bei einer Bewertung von 0,7 bzw. -20,6, für die Dialer liegen sie bei 2,5 bzw. -22,8.

Abbildung 4: Signifikante Stellen für alle Teilnehmer



Quelle: Eigene Darstellung.

Insgesamt lassen sich neun unterschiedlich lange Phasen identifizieren, die von den Teilnehmern besonders deutlich bewertet wurden. Dass die ersten fünf zugunsten *Winfried Kretschmanns* gehen und im vorderen Teil der Diskussion liegen, während sich alle Stellen zugunsten *Guido Wolfs* im hinteren Teil der Diskussion sammeln, deutet bereits an, dass der CDU-Herausforderer hier wohl stärker war als zu Beginn der Diskussion – auch wenn das Verharren der Kurve unter dem Neutralwert zeigt, dass es ihm an keinem Punkt gelang, die Stimmung zu drehen, was angesichts des Publikums allerdings nicht sonderlich überrascht. Für fünf Phasen verzeichnen Dialer und Debat-O-Meter gleichermaßen signifikante Ausschläge.

Einhellig begrüßt vom Publikum wurde hier *Winfried Kretschmanns* Warnung (Nr. 3, ab 15:23³⁸ in den Dialern, ab 15:46 im Debat-O-Meter), nach den Übergriffen von Köln das Fehlverhalten einzelner auf alle Flüchtlinge zu übertragen. Jene mit Bleibeperspektive sollten weiterhin eine Willkommenskultur erfahren, während jene zurückgeschickt würden, die keine Bleibeperspektive hätten oder die straffällig geworden seien. Ebenfalls als signifikant in beiden Messansätzen

38 Alle Zeitangaben sind relativ zum Gesprächsverlauf. Die Zeitmarke 0:00 ist dabei der Beginn der Sekunde, in der Guido Wolf erstmals auf eine Frage des Moderators Clemens Bratzler antwortet. Im Video der SWR-Mediathek ist dies die Zeitmarke 1:12.

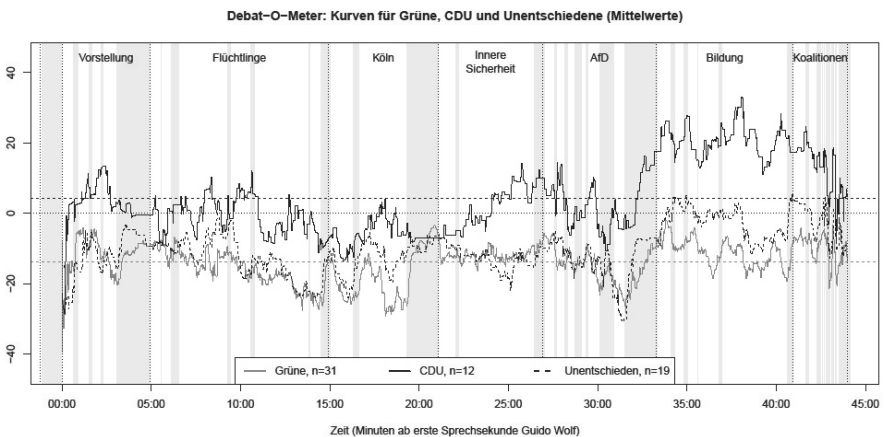
ausgewiesen und zugunsten *Kretschmanns* verbucht wird dessen Entgegnung (Nr. 4, ab 18:29 in den Dialern, ab 18:15 im Debat-O-Meter) auf den Vorwurf *Wolfs*, mit dem Zugang zu Gesundheitsversorgung für straffällige Flüchtlinge die falschen Anreize zu setzen, es zu allgemeinen menschlichen Grundsätzen gehöre, dass akut Erkrankte unabhängig vom Status ärztliche Versorgung benötigen. Der dritte, von beiden Messmethoden als signifikant ausgewiesene Abschnitt findet sich am Abschluss des Themenbereichs zur AfD (Nr. 5, ab 31:02 in den Dialern, ab 31:08 im Debat-O-Meter) und bezeichnet einen Wortwechsel, in dem *Kretschmann* sich zuerst gegen *Wolfs* Vorwurf wendet, die Grünen hätten 2011 Podiumsdiskussionen mit der Linken genauso meiden sollen wie nun im gegenwärtigen Wahlkampf mit der AfD („[nur] weil [jemand] aus dem linken Spektrum ist, ist er noch nicht irgendwie [...] verfassungsfeindlich“), worauf dieser erneut eine stärkere Beachtung linksradikaler Kräfte einfordert.

Lag *Kretschmann* bis zu dieser Stelle in der Diskussion klar vorn, relativiert sich das Bild im letzten Teil der Debatte zur Bildungspolitik etwas, wo beide Messverfahren mehrfach ein (zumindest kurzfristig) ausgewogeneres Kräfteverhältnis signalisieren. Hier finden sich auch die beiden restlichen doppelt signifikant ausgewiesenen Passagen. Zum einen kam in beiden Messungen beim Publikum die Kritik, aber auch die Bestandsgarantie und Entwicklungsperspektive *Wolfs* für die Gemeinschaftsschule gut an (Nr. 6, ab 33:43 in den Dialern, ab 33:49 im Debat-O-Meter). Punktuell Zuspruch (Nr. 7, ab 34:51 in den Dialern, ab 34:57 im Debat-O-Meter) fand aber auch seine Sprechphase, in der er kritisierte, dass der Wechsel zur Gemeinschaftsschule oft nicht aus pädagogischen, sondern vielmehr notgedrungen aus strukturpolitischen Gründen erfolge.

Die Passagen, in denen Debat-O-Meter und Dialer andere Ausschläge zeigen, sind einerseits eine dreisekündige, nur von den Dialern ausgewiesene Passage (ab 12:48, Nr. 1), in welcher *Winfried Kretschmann* der Kritik, in der Versorgung von Flüchtlingen nicht schnell genug auf Sachleistungen umzustellen, entgegenhält, dass die Umstellung ohne vorherige Ausschreibung nicht möglich sei. Inhaltlich sehr ähnlich gelagert ist die zweite ungepaarte Aussage (diesmal nur vom Debat-O-Meter erfasst, ab 14:06, Nr. 2), in der *Kretschmann* sich immer noch gegen den Vorwurf *Wolfs* wehrt, nicht schnell genug zu reagieren und mit einer bargeldlosen Karte für Flüchtlinge eine nicht praktikable Lösung anzustreben. Während beide Passagen vom Publikum zugunsten *Kretschmanns* verbucht werden, gehen die beiden folgenden Stellen auf das Konto *Wolfs*. Zum einen weist das Debat-O-Meter als auffällige Passage (ab 37:50, Nr. 8) die Phase der Diskussion aus, in der

Wolf mehr individuelle Förderung und Aufmerksamkeit für nicht-gymnasiale Bildungsformen („ein Bewusstsein, dass der Mensch nicht erst beim Abitur beginnt“) fordert. Ebenfalls zugunsten *Wolfs* verbucht wird von den Dialern eine dreisekündige Bewertungsspitze (ab 42:36, Nr. 9), als *Wolf* seine persönliche Wertschätzung für *Kretschmann* zum Ausdruck bringt („Froschkutteln-Essen“). Bei der Gesamtschau der Passagen fällt auf, dass zwei davon (Nr. 1 und Nr. 9) zeitlich relativ kurz sind und damit gegenüber den längeren, gemeinsam identifizierten Phasen nur bedingt ins Gewicht fallen.

Abbildung 5: Auswertungen für Untergruppen



Quelle: Eigene Darstellung.

Durch die geringe Zahl an CDU-Anhängern lässt sich diese Gruppe nicht separat in den Dialern ausweisen, für das Debat-O-Meter sind jedoch genügend Anhänger vorhanden, um einen cursorischen Blick in die drei wichtigsten Untergruppen (Grüne: 31, CDU: 12, Unentschiedene: 19) zu erlauben. Bereits an der mittleren Lage der beiden Kurven für prospektive Wähler der CDU (4,30) und der Grünen (-13,86) wird erneut deutlich, dass *Kretschmann* die Diskussion dominierte. Insbesondere tendiert die Kurve bei emotional besetzten Themen wie der Flüchtlingssituation oder den Übergriffen von Köln (aber auch bei der Diskussion um die AfD) deutlich zum Vorteil *Kretschmanns*, sowohl bei seinen eigenen Anhängern als auch bei denen seines Gegners. Die unentschlossenen Teilnehmer folgen hier weitgehend dem Muster der Grünen. Insgesamt ist damit die Differenz zwischen den Lagern im Publikum relativ gering, sie nimmt dann aber im Bildungsteil (und erkennbar auch gegen Ende des Themenblocks innere Sicherheit) zu. Hier tendiert die Bewertung der CDU-Anhänger deutlich stärker zu *Wolf*, aber auch Anhänger

der Grünen bewerten den Herausforderer (bei aller weiter bestehenden Ablehnung) relativ besser als in den vorhergehenden Teilen des Gesprächs. Eine ähnliche Reaktion ist auch bei den Unentschiedenen zu sehen, deren mittlere Bewertung hier stärker von jener der Grünen abweicht und sich auf *Wolf* zubewegt, ohne jedoch mehr als punktuell über die neutrale Mitte zu wechseln. Durch die deutliche Tendenz der CDU-Anhänger zu *Wolf*, in Verbindung mit der vorhandenen, aber weit weniger klaren Bewegung weg vom Vorteil für *Kretschmann* durch die Anhänger der Grünen, wächst im Aggregat die Differenz zwischen beiden Kurven, sodass die Bildungsdiskussion der Teil der Debatte ist, in dem die Meinungen im Publikum am deutlichsten auseinandergehen.

2. Die Elefantenrunde am 10. März 2016

Der zweite Versuch fand zur *Elefantenrunde* des SWR im Audimax der Universität Freiburg statt. Insgesamt folgten rund 120 Personen der offenen Einladung, die Diskussion live zu sehen. Da die Debatte sechs Personen umfasste – *Winfried Kretschmann* (Grüne), *Nils Schmid* (SPD), *Guido Wolf* (CDU), *Hans-Ulrich Rühle* (FDP), *Bernd Riexinger* (Linke) und *Jörg Meuthen* (AfD) –, war ein Parallelversuch mit Dialern nicht möglich, weshalb alle Teilnehmer ihre Bewertung über das Debat-O-Meter abgaben. Als Messanweisung sollten die Teilnehmer erneut ihren globalen Eindruck wiedergeben, diesmal allerdings von den einzelnen Kandidaten, die sie punktuell (d.h. nicht kontinuierlich) durch Drücken eines „+“- oder „-“-Buttons bewerten konnten. Auch diese Teilnehmer wurden zu einer Vor- und Nachbefragung gebeten. Insgesamt nahmen 74 Personen an der Vorbefragung teil, für 59 Personen liegen Angaben aus Vor- und Nachbefragung vor. Im Debat-O-Meter sind insgesamt 91 Kurven erfasst, allerdings wurden acht Kurven entfernt, da die Teilnehmer die Bewertung während der Diskussion abbrachen, sodass insgesamt 83 Kurven in die Analyse einfließen.

Ähnlich der Veranstaltung zum Duell *Kretschmann/Wolf* bestand das Publikum gemäß der Vorbefragung zu über 55 % aus Studierenden (19 % machten keine Angabe). Zwei Drittel waren Männer, 43 % wollten die Grünen wählen bzw. hatten dies bereits getan (21 % der Teilnehmer hatten per Briefwahl gewählt), 14 % nannten die SPD, 12 % die Linke, 4 % die AfD, jeweils 1 % nannte CDU und FDP, 24 % waren noch unentschieden. Über 74 % würden *Winfried Kretschmann* als Ministerpräsidenten direkt wählen, 8 % *Bernd Riexinger*, 11 % präferierten keinen der Kandidaten, niemand nannte *Guido Wolf*.

Anders als beim Duell kann wegen der Bewertung durch Buttons zur Aggregation nicht direkt über alle RTR-Kurven gemittelt werden. Um dennoch eine Aussage

auf Kandidatenebene zu erhalten, wurde daher für jeden Kandidaten sekundenweise die Anzahl eingehender positiver und negativer Bewertungen saldiert. Um festzustellen, ob dieser Saldo zu einer gegebenen Sekunde stark von der mittleren Publikumsbewertung abweicht, wurde der Durchschnitt über alle Zeitpunkte berechnet, in denen der jeweilige Kandidat Votes erhielt und keiner der Moderatoren sprach. Sekunden, in denen der Saldo mehr als 1,96 Standardabweichungen von diesem Mittelwert abweicht, werden als hervorgehoben angesehen. Da sich, anders als am 14. Januar, kaum Anhänger von CDU und FDP im Publikum befanden, werden im Folgenden keine Untergruppen betrachtet.

Anders als bei der Slider-basierten Messung treten durch die punktuelle Vergabe von Votes (vgl. die Oberflächen in *Abb. 1*) deutlich die einzelnen Sprechphasen der Kandidaten hervor, ein Hinweis darauf, dass die Teilnehmer vor allem auf die verbale Ebene abgestellt haben, auch wenn dies in der Messanweisung nicht ausdrücklich so angegeben wurde. Mittelt man die Salden der Kandidaten (unter Ausschluss nicht bewerteter Zeiten und Moderationsphasen), zeigt sich deutlich, dass das linke Lager beim Publikum vorne lag: Amtsinhaber *Winfried Kretschmann* ist im Schnitt mit 2,35 Votes im Plus, wird aber sowohl von *Nils Schmid* (+2,59) als auch von *Bernd Riexinger* (+3,27) überflügelt. Deutliche Ablehnung erfährt das bürgerliche Lager, wobei *Hans-Ulrich Rülke* (-1,25) noch vor *Guido Wolf* (-1,60) liegt. Von AfD-Spitzenkandidat *Jörg Meuthen* hatte das Publikum mit Abstand den negativsten Eindruck: Wird er bewertet, hat er pro Sekunde im Saldo durchschnittlich 3,05 Zuschauer gegen sich.

Thematisch stand die Debatte einerseits im Zeichen der Flüchtlingskrise, andererseits hatte das Erstarken der AfD im Vorfeld der Wahl und der Streit um ihre Teilnahme die Erwartung geschürt, dass es zu einer grundsätzlichen Auseinandersetzung mit der Partei kommen würde, sodass es sinnvoll erscheint, beide Bereiche näher zu beleuchten. Im Hinblick auf die Flüchtlingskrise ernteten einerseits Forderungen einer europäischen bzw. Warnungen vor nationalen Lösungen (vorgelesen von *Kretschmann* (4:07³⁹), *Wolf* (4:33) und *Schmid* (7:42)) deutliche Zustimmung des Publikums, ebenso wie die Verbindung der Thematik mit menschlichen Grundwerten einerseits sowie von Fluchtursachen und deutschen Waffenexporten andererseits durch *Bernd Riexinger* (12:20 bzw. 12:45). Kritik an der Flüchtlingspolitik der Kanzlerin (z. B. *Hans-Ulrich Rülkes* Aussage gleich zu Be-

39 Wie beim Duell Kretschmann/Wolf ist die Marke 0:00 für die angegebenen Zeiten die Sekunde, in der der erste Kandidat (hier Hans-Ulrich Rülke) erstmals auf eine Frage antwortet.

ginn, sie habe „eine halbe Völkerwanderung“ ausgelöst, 0:25) bzw. an der Unterstützungsleistung der Grünen für ihre Politik (*Wolf*, 5:32) trugen den Kandidaten dagegen signifikant negative Salden ein. Die stärksten Reaktionen in diesem Teil der Diskussion produzierten jedoch Auseinandersetzungen mit der AfD: Zum einen erhielten *Jörg Meuthens* Diagnose eines „Kontrollverlust[s]“ in der Flüchtlingskrise (8:15) und seine Forderung eines Grenzzauns als „ultima ratio“ (8:58) lang anhaltende negative Voten aus dem Publikum, zum anderen traf *Nils Schmid*s Gegenrede, wonach die AfD eine „menschenverachtende Ideologie“ vertrete, auf starke länger anhaltende Zustimmung (11:25).

Ähnlich dem Muster im Themenbereich zur Flüchtlingskrise generierten im anschließenden Themenblock zur AfD erneut Angriffe *Schmid*s und *Wolfs* auf die Partei die deutlichste Zustimmung des Publikums. Während *Kretschmann*s Vorwurf an *Meuthen*, keine Verantwortung zu übernehmen, zwar leichten Zuspruch erhielt, bekam der Ministerpräsident später negative Salden bei seiner Verteidigung der Rückführungszahlen sowie seiner Beschreibung der Erstaufnahmesituation (21:21 und 25:26). Stark und lang anhaltend positiv bewertet wurden hingegen die Vorwürfe *Wolfs*, der AfD-Spitzenkandidat distanzieren sich nicht von den Aussagen seiner eigenen Kandidaten (16:41), sowie *Schmid*s Statement „Anständige Leute wählen keine Rassisten“ (21:38) – und damit erneut direkte Angriffe auf die AfD. Negativster Punkt für *Jörg Meuthen* hingegen war seine Aussage, dass die AfD „klare Lösungsvorschläge“ biete (18:21).

Zu den weiteren auffälligsten Stellen der Diskussion zählten u. a. für Ministerpräsident *Winfried Kretschmann* seine – vom Publikum positiv bewertete – Auseinandersetzung um die Kennzeichnungspflicht für Polizisten mit *Guido Wolf*, für den die Passage zugleich den am schlechtesten bewerteten Punkt der Diskussion darstellt (45:20). Am meisten Zuspruch mit einem Maximalsaldo von 21 positiven Bewertungen erhielt *Kretschmann* hingegen für seine Feststellung im Themenfeld Wirtschaft und Infrastruktur, dass Straßenneubau ohne Bestandserhalt ökonomisch unvernünftig sei (68:00). Stärkster Moment seines Koalitionspartners *Nils Schmid* war aus Sicht des Publikums eine Maßregelung *Hans-Ulrich Rülkes* im Gespräch um die Polizistenausbildung („Nehmen Sie mal den Finger da weg“, 37:52), am schlechtesten bewertet wurde hingegen die Aussage am Ende der Diskussion, dass eine Stimme für die Linke eine für Grün-Rot verlorene Stimme sei (84:07). Auf Seiten der AfD ist die am schlechtesten bewertete Passage der Diskussion die Aussage *Jörg Meuthens*, dass für die „unzureichende“ Verfolgung ethnisch und religiös motivierter Straftaten durch die Polizei keine Belege vorlägen (42:53).

V. Die Debattenwirkung

Die Aufmerksamkeit, die TV-Duellen zu Teil wird, hängt nicht zuletzt daran, dass man ihnen zuschreibt, die Einstellungen von Wählern verändern zu können. Dieser Abschnitt versucht, in groben Zügen die Effekte der Debatte auf das Publikum nachzuzeichnen. Dabei konzentriert sich die Analyse auf die Gegenüberstellung von Vor- und Nachbefragung, es werden also jene Teilnehmer betrachtet, die in beiden Fällen Auskunft gegeben haben.

1. Das TV-Duell Winfried Kretschmann/Guido Wolf am 14. Januar 2016

Insgesamt haben 184 Personen sowohl an der Vor- wie Nachbefragung teilgenommen. Vergleicht man, wer zu Beginn als Sieger der Diskussion erwartet wurde und wer danach als tatsächlicher Gewinner galt, so spiegelt der Befund einerseits klar die Zusammensetzung des Publikums wider, andererseits reflektiert er aber auch das Kräfteverhältnis nach der RTR-Messung: Für 62,4% war *Winfried Kretschmann* bereits vor der Debatte Favorit, gleichwohl erwarteten aber auch 28,7% ein Unentschieden; nur 8,8% erwarteten, dass *Guido Wolf* die Diskussion für sich entscheiden würde. In der Nachbefragung erhöhte sich sein Anteil kaum auf 9,4%, *Kretschmann* wurde nun jedoch von 80,0% als Sieger angesehen. Der gleiche Befund zeigt sich, wenn man nach dem individuellen Abschneiden prüft: Auf einer Skala von 1 („sehr schlecht“) bis 5 („sehr gut“) übertraf *Kretschmann* sowohl die Erwartungen der Anhänger des Regierungslagers (vor der Debatte: 3,9, danach: 4,2) als auch der Unentschiedenen (3,6 auf 4,0) und erfüllte die Erwartungen der Anhänger von FDP und CDU (3,3 vor und nach der Debatte). *Guido Wolf* hingegen konnte nur in seinem eigenen Lager punkten (3,1 auf 3,5), Unentschiedene (3,1 auf 2,4) und Regierungsanhänger (2,8 auf 2,2) attestierten ihm jedoch ein schlechtes Abschneiden.

Etwas anders stellt sich die Lage im Hinblick auf *Guido Wolfs* Problem zu Beginn des Wahlkampfs dar, in der Wählerschaft nicht genug bekannt zu sein. Hatten in der Vorbefragung noch 34,8% angegeben, dass ihnen der Kandidat für ein Urteil zu unbekannt sei, trauten sich hinterher nur noch 2,2% keine Bewertung mehr zu (*Kretschmann*: 6,1% bzw. 1,7%). Im Spiegel der Bewertungen auf einer Skala von -5 bis +5 ist dieser Erfolg jedoch sehr fragwürdig: Während *Wolf* sich im bürgerlichen Lager zwar verbessern konnte (1,1 auf 1,5), fiel seine mittlere Bewertung bei den Regierungsanhängern von -2,5 auf -2,6 und unter den Unentschiedenen von -0,9 auf -1,6. Anders das Bild für *Kretschmann*: Er konnte in allen drei Lagern klar punkten (Regierungsanhänger: 3,2 auf 3,7; Unentschiedene: 2,2 auf 2,4; Anhänger von CDU und FDP: 1,5 auf 2,2). Dementsprechend verwundert es

kaum, dass vier Fünftel (80,6% bzw. 25 Personen), die ihre Präferenz für eine Direktwahl des Ministerpräsidenten änderten, zu *Kretschmann* wechselten (wobei 23 aus dem Lager der zuvor Unentschiedenen stammten).

Tabelle 1: Veränderung der Wahlabsicht nach der Debatte am 14.01.2016

		Nach der Debatte								
		AfD	Andere	CDU	FDP	Grüne	Linke	Unentsch.	Nichtwahl	SPD
Vor der Debatte	AfD	1		1						
	Andere		6					1		
	CDU			12				2		
	FDP				6	1		1	1	
	Grüne					71	1	2		
	Linke					1	10	2		
	Unentsch.			4	3	16		18		
	Nichtwahl					1				
SPD						3			11	

Quelle: Eigene Darstellung.

Wie ist der Effekt der Diskussion auf die Wahlentscheidung einzuschätzen? *Tab. 1* zeigt die Veränderungen der Wahlentscheidungen durch das Duell. Direkt sichtbar ist, dass vor allem die Grünen profitieren konnten, entschlossen sich doch 22 Personen, die zuvor noch eine andere Entscheidung angegeben hatten, für diese Partei (während nur drei das Lager verließen). Die CDU hingegen konnte fünf Wähler hinzugewinnen, verlor zugleich aber auch zwei, die sich ihrer Entscheidung nach der Debatte nun nicht mehr sicher waren. Von jenen 16 Personen, die in der Nachbefragung die Grünen wählen wollten, hatten vier eine entsprechende Parteiidentifikation, sodass sie als mobilisierte Wähler gelten können, der Rest war anderweitig gebunden (zwei machten keine Angaben), sodass Konversionen der Wahlentscheidung überwiegen. Für das bürgerliche Lager hatten alle, die nach dem Duell CDU oder FDP wählen wollten, bis auf zwei eine entsprechende Parteibindung, sodass hier klar die Mobilisierung überwiegt. Auffällig ist auch die relativ starke Demobilisierung, da immerhin acht Personen ihre ursprüngliche Wahlentscheidung aufgaben und nach dem Duell nicht mehr sicher waren, wen sie wählen wollten.

2. Die Elefantenrunde am 10. März 2016

Insgesamt haben 59 Personen an Vor- und Nachbefragung teilgenommen, sodass wir die Veränderungen ihrer Einschätzungen durch die Diskussion verfolgen können. Bereits der erste Vergleich zwischen dem erwarteten und dem tatsächlichen Gewinner der Debatte zeigt hier relativ deutliche Verschiebungen, die erneut dem Bild der RTR-Messung entsprechen. Ging *Winfried Kretschmann* mit 44,1 % als klarer Favorit ins Rennen (40,7 % erwarteten keinen klaren Sieger, 11,8 % waren sich nicht sicher), fanden nach der Diskussion nur noch 35,6 %, dass er tatsächlich am besten abgeschnitten hatte; dafür nannten nun 33,9 % *Nils Schmid* bzw. 8,5 % *Bernd Riexinger* (beide Politiker waren in der Vorbefragung nicht genannt worden). Keinen klaren Sieger wollten 13,6 % ausmachen. In ähnlicher Form finden sich diese Bewertungen auch auf Ebene der einzelnen Kandidaten. Auf einer Skala von 1 („sehr schlecht“) bis 5 („sehr gut“) blieb das Abschneiden *Kretschmanns* in der Debatte geringfügig hinter den Erwartungen des Publikums zurück (durchschnittliche Erwartung vor dem Duell 4,22, Einschätzung danach 4,09⁴⁰). *Nils Schmid* schnitt hingegen deutlich besser ab als erwartet (3,13 auf 4,02), wie auch *Bernd Riexinger* (3,15 auf 3,62). *Guido Wolf* (2,57 auf 2,72) lag zumindest etwas über den Erwartungen des Publikums, *Hans-Ulrich Rülke* mit 2,90 vs. 2,65 dagegen etwas darunter. Deutlich verlor *Jörg Meuthen*, der mit 3,24 gegenüber 2,08 klar schlechter abschnitt als erwartet.

Ähnlich der Diskussion am 14. Januar sammelte das Publikum auch hier Kenntnis über die Kandidaten. Hatten von den 59 Teilnehmern zuvor jeweils 19 angegeben, nicht genug Kenntnis über *Bernd Riexinger* und *Hans-Ulrich Rülke* zu haben, um eine Einschätzung abgeben zu können (zehn wählten diese Option für *Jörg Meuthen*), wurde nach der Diskussion diese Antwortoption nie öfter als zwei Mal gewählt. Auf dem Skalometer von -5 bis +5 verschob sich die Bewertung von *Winfried Kretschmann* kaum (2,76 vor der Debatte, 2,97 danach), wie auch die von *Jörg Meuthen* (-3,50 vor und nach der Diskussion). *Guido Wolf* hingegen gewann hinzu (von -1,96 auf -1,03), wie auch *Bernd Riexinger* (0,76 auf 1,68). Auffallend sind jedoch die Veränderungen für *Nils Schmid*, der – hier aufgeschlüsselt nach Anhängern von Grünen/SPD bzw. Unentschlossenen als größte Gruppen – im eigenen Lager zwar punkten konnte (von 1,78 auf 2,63 bei prospektiven Wählern der grün-roten Koalition), noch besser aber bei den Unentschlossenen ankam

40 Aufgrund der Zusammensetzung des Publikums konnten nur Anhänger des grün-roten Regierungslagers und Unentschiedene als Untergruppen untersucht werden. Da sich die Tendenzen in beiden Gruppen nicht sonderlich unterschieden, wird hier nur der Wert für das gesamte Publikum ausgewiesen.

(von -0,60 auf 1,60); ähnlich die Tendenz für *Hans-Ulrich Rülke* (von -1,95 im grün-roten Lager auf -1,35 und von -1,75 unter den Unentschlossenen auf -0,73).

Tabelle 2: Veränderung der Wahlabsicht nach der Elefantenrunde am 10.03.2016

		Nach der Debatte								
		AfD	Andere	CDU	FDP	Grüne	Linke	Unentsch.	Nichtwahl	SPD
Vor der Debatte	AfD	1						1		
	Andere									
	CDU			1						
	FDP					1				
	Grüne			1		17	1			3
	Linke						7			
	Unentsch.					4	1	11		
	Nichtwahl									
	SPD					1				8

Quelle: Eigene Darstellung.

Tab. 2 zeigt die Veränderungen der Wahlpräferenz durch die Diskussion. Im Aggregat sind nur relativ geringe Verschiebungen sichtbar. Während die Grünen fünf Wähler verloren (drei davon an die SPD), konnten sie sechs dazu gewinnen, die meisten aus den Reihen der Unentschlossenen. Aufgrund der geringen Fallzahlen lässt sich aber leider nur spekulieren, ob der Abgang grüner Wähler zur SPD mit dem Abschneiden ihres Spitzenkandidaten begründet werden kann.

VI. Zusammenfassung

Mit dem Debat-O-Meter liegt nun der voll funktionsfähige Prototyp eines RTR-Messinstruments vor, das auf Smartphones, Tablets, aber auch auf jedem anderen mobilen oder nicht-mobilen PC oder Notebook mit Internetverbindung funktioniert und das bereits mehrfach erfolgreich für Messungen politischer Debatten eingesetzt wurde. Wichtigster Vorteil gegenüber bestehenden Methoden ist, dass Forscher keine spezialisierte Hardware (*Drehregler*) mehr benötigen, um eine solche Echtzeitmessung einer Debatte durchzuführen. Neben der damit verbundenen Kostenersparnis ist das System vor allem deshalb interessant, weil internetfähige Geräte heutzutage in fast jedem Haushalt anzutreffen sind. Dies eröffnet die Perspektive, zukünftig Echtzeitmessungen außerhalb des Hörsaals durchführen zu können und die potenziellen Beschränkungen dieses Settings zu umgehen. Damit

kann ein weiteres Problem bisheriger Messungen gelöst werden: die mangelnde Repräsentativität. Alle bisherigen Messungen mit Drehreglern, aber auch die des Debat-O-Meters, krankten an der mangelnden Repräsentativität der Teilnehmer. Mit einer hohen Zahl an Teilnehmenden wird es möglich sein, über Quoten eine deutlich bessere Repräsentativität herzustellen, insbesondere auch räumlich.

Gegenstand dieses Beitrags war es, anhand der bisher gesammelten Daten drei Fragen nachzugehen, einerseits zur methodischen Herausforderung eines neuen Messinstruments sowie andererseits zur Landtagswahl Baden-Württemberg 2016: Inwieweit ist das Debat-O-Meter mit herkömmlichen RTR-Messgeräten vergleichbar? Wie haben die Zuschauer das TV-Duell sowie die Elefantenrunde zur Landtagswahl wahrgenommen? Und: Welche Effekte hatte die Debatte bei ihnen?

Die bisherige Antwort auf die erste Frage ist im Grundsatz positiv: Die RTR-Kurven zweier randomisierter Testgruppen zum TV-Duell am 14. Januar 2016 waren in der Lage praktisch identisch und wiesen neben einer ähnlichen Verteilungsform und trotz etwas geringerer Streuung auf Seiten des Debat-O-Meters eine hohe Interkorrelation auf. Auch in der Bedienung waren für die Teilnehmer keine gravierenden Unterschiede auszumachen. Zwar scheinen herkömmliche Dialer geringfügig einfacher zu bedienen, aber in der Summe fallen die Differenzen kaum ins Gewicht. Und auch inhaltlich kommt man auf Basis beider Systeme zu weitgehend gleichen Schlüssen. Zwar stellen die Ergebnisse dieses Beitrags noch keinen abschließenden Beleg für die Äquivalenz der Systeme dar, sie stimmen aber optimistisch, dass das Debat-O-Meter bestehenden Systemen ebenbürtig ist.

Mit Blick auf die zweite Frage zeigte sich für die beiden Debatten eine recht unterschiedliche Situation: War Ministerpräsident *Kretschmann* klar der Sieger des ersten Aufeinandertreffens, wurde sein Abschneiden bei der Elefantenrunde zwar weiter positiv, aber doch verhaltener bewertet. Deutlich besser konnte hingegen sein Koalitionspartner *Nils Schmid* abschneiden, der für seine konfrontative Diskussionsführung vom Publikum klaren Zuspruch erhielt. Für Herausforderer *Guido Wolf* waren die Debatten ein zweischneidiges Schwert. Zwar konnte er in der ersten Diskussion seine damalige relative Unbekanntheit überwinden und auch im Bildungsbereich das Publikum teilweise für sich gewinnen, gleichwohl gelang es ihm nie, die Situation für sich zu drehen. Etwas besser verlief dagegen für ihn die zweite Diskussion, wo vor allem sein Auftreten gegen die AfD beim Publikum Anklang fand.

Mit Blick auf die dritte Frage zeigen unsere Daten im Einklang mit der Literatur, dass Debatten den Eindruck von Politikern mitunter deutlich verändern können.

Neben der bereits erwähnten Konkretisierung der Einschätzungen wurden vor allem für die Elefantenrunde starke Veränderungen der Bewertung sichtbar, die teilweise auch über Untergruppen variierten. Als Beispiel ist hier erneut *Nils Schmid*s Auftritt in der Elefantenrunde zu nennen, der bei den Unentschiedenen sogar noch besser ankam als im eigenen Lager. Aber auch *Guido Wolf*s Auftritt im TV-Duell vom 14. Januar fällt in diese Kategorie, erlitt er doch deutliche Verluste in der Bewertung bei den Unentschiedenen.

In seiner bisherigen Fassung kann das Debat-O-Meter bereits sehr flexibel in klassischen RTR-Laborsettings benutzt werden. Für die nähere Zukunft sollten vor allem zwei Vorteile weiterentwickelt werden, die aus der Virtualisierung des Messinstruments entstehen: Einerseits sollte konsequent die Möglichkeit geschaffen werden, das Debat-O-Meter für die Anwendung im privaten Bereich zu entwickeln, um die bisherige Beschränkung klassischer RTR-Designs auf Laborsettings zu beheben. Andererseits sollte auch die Flexibilität einer Bildschirmoberfläche stärker genutzt werden, bietet sich hier doch die Möglichkeit, das volle Repertoire des Designs grafischer Nutzeroberflächen dazu einzusetzen, sozialwissenschaftliche Messkonzepte besser zu erfassen.