Time Dynamics and Stability of Political Identity and Political Communication

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Jacob A. Long, M.A.

Graduate Program in Communication

The Ohio State University

2020

Dissertation Committee

William P. Eveland, Jr., Adviser

Gerald M. Kosicki

Michael D. Slater

Christopher R. Browning

Abstract

Several of the most important constructs in political communication research appear to be quite stable over time: media use, political discussion, and partisanship. If these constructs change only very rarely, it would cast doubt on theories presuming that they affect one another. On the other hand, it may be the case that they are only conditionally stable, such that if some factor or factors were altered, change in these constructs would be expected. This project explores the possibility that communication can both promote and undermine consistency in identity, a little-appreciated type of effect in communication research. The apparent stability of attitudes and behaviors has figured prominently in several prominent scholarly findings and debates in the discipline. I describe some ambiguity in how researchers define stability and provide a clear framework and definition for stability and communication effects. From there, I demonstrate how stability can be analyzed in the context of quantitative research designs. My approach to modeling stability can seamlessly integrate both the individual-level predictors of stability while still allowing for the discovery of the more familiar type of media effect that change the mean level of a variable. Furthermore, a technical analysis of the statistical evidence claiming to demonstrate high stability of communication and identity reveals that such claims are likely overstated.

Using insights from social identity research, the project shows how there may be more variability than meets the eye when it comes to partisan identity. Rather than categorical shifts between Republican and Democrat, this variability comes in the form of changes in the strength of the identity as well as the impact of the identity on one's self-esteem. It was expected that identity motivates communication, yet communication also affects identity. Rather than cause

extreme identities and excessive communication, however, people generally achieve an equilibrium in which their communication is just sufficient to maintain their current level of affiliation. Although congenial political communication may bolster the strength of partisan identity, I argue that those with strong identities require some level of such communications just to resist an inherent pull toward moderation.

To explore these relationships, an intensive longitudinal design with daily surveys of the same group of respondents over a 3-week timespan was used. Having granular information on communication and political identity made it possible to capture the real variation in these constructs that occurs on a day-to-day basis, especially communication behaviors that are constrained by the routines of daily life like work, school, and the weekend. Results suggest that congenial political communication indeed increase the strength of one's partisan identity. Although some forms of political communication, especially discussion, can destabilize identity, the evidence suggests that in-party sources promote identity stability. The results also suggest an important role of exposure to differing views as potentially important for moderating and destabilizing identities.

Acknowledgments

I would like to begin by acknowledging my adviser, Dr. Chip Eveland. When I visited Ohio State as a prospective student, I had a meeting with Chip. I expressed some uncertainty about whether Ohio State was right for someone like me, whose background was in film studies and political philosophy and had no statistical training since high school. Rather than simply sell the program, Chip seriously entertained the idea that Ohio State could be the wrong place for me and helped me think through the decision I had to make. Looking back, I find this interaction typical of his advising style: He is generous with his time and energy, he is straightforward about what he thinks, and he is not eager to impose his views onto others. As I tried to find my place in this little world of communication research, Chip gave me space to figure things out with just the right amount of nudging to make sure I kept my eyes on the prize. I really enjoyed my time in graduate school and there is no doubt that the positive advising relationship I had was a necessary condition for that.

I also had the pleasure of getting to know Dr. Jerry Kosicki in multiple courses, as his research assistant, and in his capacity as committee member. He has been an invaluable source of scientific and institutional knowledge as well as a great sounding board and mentor. I also want to recognize Dr. Mike Slater, who allowed me to take a leadership role in one of my first research collaborations. He always took my ideas seriously, which included asking tough questions, and I am especially grateful that he stuck with me as he transitioned into his new role as Director. And a big thanks to Dr. Chris Browning for joining the committee at such a late stage — my meeting in his office was one of the last times I was ever on campus.

Besides the members of my committee, I would like to recognize Dr. Kelly Garrett as someone who I have looked up to and helped to teach me to think like a scientist. Another person who I have considered a role model and a particularly valuable source of socialization into the discipline is Dr. Jesse Fox. Likewise, if I had not crossed paths with Dr. Robert Bond, I am not sure whether I would ever have picked up my interest and skills in computational social science. All the aforementioned people also always stood out to me in seminars, colloquia, and so on as particularly insightful and constructive participants who greatly enhance the intellectual atmosphere of the School of Communication. When you know these people are around to evaluate your work, you always bring your A-game.

Prior to my arrival in Columbus, I had two important mentors at Rhodes College who were essential to me finding this career path. One is Dr. Rashna Richards, who introduced me to the study of film and was the first to encourage my graduate school aspirations. The other is Dr. Amy Jasperson, who pointed out to me that there was a discipline that fused my interests in politics and media and then led me through a crash course in it.

I also must thank Alex Sterne for applying and getting admitted to Ohio State, since I would have neither applied nor enrolled had she not done so. She then put up with me for six years of graduate school, making them the best time of my life (so far). The sacrifices and efforts of my parents, Dave and Peg Long, were of course essential both to my arrival at Ohio State and my continued flourishing once I got there. I thank my brothers Chad and Michael for, among many other things, including their forever-student little brother involved in often-expensive family activities. Thanks to my sister Kim for bringing my mother so much happiness and being

a part of our unconventional family. I thank my brother Josh for telling me not to waste time doing things I don’t enjoy.

Vita

2010.....Pontiac Township High School, Pontiac, IL

2014.....B.A. with Honors, Political Science; Rhodes College

2019.....M.A., Communication; The Ohio State University

Publications

Long, J. A., & Eveland, W. P., Jr. (2020). Entertainment use and political ideology: Linking worldviews to media content. *Communication Research*. OnlineFirst. https://doi.org/10.1177/0093650218791011

Garrett, R. K., **Long, J. A.**, & Jeong, M. S. (2019). From partisan media to misperception: Affective polarization as mediator. *Journal of Communication*, *69*, 490–517. https://doi.org/10.1093/joc/jqz028

Long, J. A., Eveland, W. P., Jr., & Slater, M. D. (2019). Partisan media selectivity and partisan identity threat: the role of social and geographic context. *Mass Communication and Society*, *22*, 145–170. https://doi.org/10.1080/15205436.2018.1504302

Fields of Study

Major Field: Communication

Graduate Interdisciplinary Specialization in Survey Research

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

This dissertation takes as a point of departure the observation that several of the most important constructs in political communication research appear to be quite stable over time: media use (Scharkow, 2019), quantity and qualities of political discussion (Bello & Rolfe, 2014; Sokhey, Baker, & Djupe, 2015), and party identity (Green & Yoon, 2002; Johnston, 2006). If this is intrinsic to the constructs — people just do not change their political identities or their communication behaviors — it would cast doubt on theories presuming that they affect one another. On the other hand, it may be the case that they are only conditionally stable, such that if some factor or factors were altered, change in these constructs would be expected. The goal of this dissertation is to examine this possibility. I hypothesize that the fact these constructs have trait-like stability is not coincidental, but instead is because they are causally related to one another. In particular, I focus on how communication can promote consistency in identity, a little-appreciated type of effect in communication research.

I begin by discussing the role of stability in communication research, focusing on its recurring role in several prominent scholarly findings and debates. The consistency of partisanship has played a major role in mass communication research since the era of Paul Lazarsfeld, causing many to feel media effects were too small to be interesting during the field's early years (Berelson, 1959; Klapper, 1960). The debates that followed (Lang & Lang, 2006; Schramm, 1959) are echoed today as the discipline considers whether communication and partisanship are so tightly linked — thereby promoting reinforcement rather than change — that a new paradigm is needed (Bennett & Iyengar, 2008; Holbert, Garrett, & Gleason, 2010). I

describe some ambiguity in how researchers define stability and the closely related concept, reinforcement, ultimately offering a conceptualization of stability as consistency.

I conceptualize partisan identity as a social identity, applying insights from social psychological research on identity to this aspect of politics (Greene, 2004; Iyengar, Sood, & Lelkes, 2012). Recent research has suggested partisan identity, moreso than ideology or issue attitudes, drives political behavior (Huddy, Mason, & Aarøe, 2015). Using these insights from social identity research, I discuss the ways there may be more variability than meets the eye when it comes to partisan identity. Rather than categorical shifts between Republican and Democrat, the expected variability comes in the form of changes in the strength of the identity as well as the impact of the identity on one's self-esteem.

I then turn to the plausibility of my hypothesized connection between communication and partisan identity. It is well-documented, at least in the United States[1], that partisans show preferences for like-minded media (e.g., Iyengar & Hahn, 2009) and like-minded people for political discussion (e.g., Mutz, 2006), although this is rarely to the complete exclusion of disagreeable media (Garrett, 2009) or people (Eveland, Appiah, & Beck, 2018; Huckfeldt, Johnson, & Sprague, 2004). The research establishing these phenomena, however, rarely engage with social identity besides engaging in the not unreasonable assumption that co-partisans share many attitudes. With this in mind, I consider the applicability of the theoretical mechanisms specified by selective exposure and related theories. I ultimately conclude there are good reasons

[1] Fletcher and Jenkins (2019) review research in Europe, finding cross-national variation in the extent of selective exposure.

to believe communication can be and is motivated by social identity, but there is surprisingly little research engaging this question directly. The Reinforcing Spirals Model (RSM; Slater, 2007, 2015) provides a theoretical framework from which it is possible to derive predictions regarding the interdependence of partisan identity and communication. In short, it is expected that identity strength motivates identity-relevant — and identity-affirming — communication, which in turn promotes identity stability. I propose testing a simplified model compared to the RSM which involves the assumption that identity strength recedes in the absence of the kind of affirmation provided by communication.

Next, I review what is known about the stability of these constructs. As mentioned, each of partisan identity, media use, and the composition of political discussion networks have been shown to be quite stable. They are indisputably trait-like constructs in that the relative level for an individual at one point in time is highly predictive of their relative level months and years later. Exceptionally high stability, however, would at minimum make compelling research designs intractable due to a lack of within-person variability that can be used to analyze causal processes. More problematic is that such high stability would cast doubt on the very idea that there are causal processes to be studied at all. The research on media stability (Lee, Hornik, & Hennessy, 2008; Scharkow, 2019) suggests it indeed may be too stable to study — at least over the course of months as in a typical political campaign study. There is also similarly foreboding work on the stability of partisanship in the United States (Green & Palmquist, 1990), although in this case I will argue the commonly used measure is flawed when treated as continuous and likely to inflate estimates of measurement error (Clarke & McCutcheon, 2009). The research is less clear in the case of political discussion, mainly due to there being no efforts of which I am

aware to quantify it in a way that is directly comparable to approaches used for media and identity.

Furthermore, I raise questions about some of the published claims of high stability. I describe the methods most commonly used for the quantification of stability in technical detail, showing how on one hand it may not express the most interesting sort of stability and in any case involves rarely-explored assumptions that are reasonably likely to be violated for these constructs. From there, I pivot to a detailed description of how stability can be examined at the level of the individual person. My approach to modeling stability can seamlessly integrate both the individual-level predictors of stability and of more enduring changes that can be conceptualized as developmental and are akin to the typical sort of media effects.

To test my hypotheses in a targeted way, I will use an intensive longitudinal design[2] (Bolger & Laurenceau, 2013) with daily surveys of the same group of respondents over a 3-week timespan. This design is chosen for several reasons. As I will review, there is some evidence that the constructs of interest are quite stable over the periods of time covered by more common panel designs. I will discuss how this evidence may be flawed if (among other things) the true change in the constructs follows a pattern of variance around a mean rather than trending up or down. I cannot demonstrate that this is the case without more measures for each person. Furthermore, by shortening the frame of reference for the questionnaire measures (i.e., the last

[2] I avoid the term "experience sampling" because it implies that participants will respond when prompted by specific events. Instead, the proposed design is more like a long panel survey with very short lags.

day), the cognitive burden on respondents to provide accurate responses about their behaviors is reduced. Normal day-to-day variation in the constructs of interest are unlikely to be captured when questionnaires inquire about "regular" behaviors or refer to timeframes on the scale of weeks or months. The longitudinal design allows me to model stability and change in identity as well as communication behaviors, including the interdependence thereof.

Chapter 2. Stability and Reinforcement in Communication Research

The claim that mediated communication has limited influence on attitudes and behavior has played a prominent role in the history of communication research. Lazarsfeld, Berelson, and Gaudet (1948), for instance, claim to have found "no overt effect on vote behavior at all" (p. 87) in their trailblazing Erie County study. Lazarsfeld et al. (1944/1948) had a nuanced message on this topic, but the received history of the study and its contemporaries is one that emphasizes a lack of media effects (e.g., Klapper, 1960). Although some have reconsidered whether this was an accurate characterization of the results of media research in the 1940s and 1950s (e.g., Gitlin, 1978), given the existence of contrary findings (e.g., Lang & Lang, 1953) and modern reanalyses of the data, more important than the findings themselves is the remembered history[3]. Near the end of this era of supposed minimal effects, some of the field's brightest minds cast doubt on the continued usefulness of (mass) communication research (Berelson, 1959), something even those who were not so pessimistic struggled against (Lang & Lang, 2006).

Of course, the field moved on despite the doubters and eventually entertained theories suggesting very strong effects of media (e.g., Gerbner & Gross, 1976). Thanks in no small part to some true believers in the research enterprise in Berelson's (1959) era, what Schramm (1959) called "a field, not a discipline" (p. 8) looks much more like a discipline today with great popularity among undergraduate students and broad institutionalization in major universities. One positive outcome of the one-time acceptance of minimal media effects was the Columbia

[3] To borrow a phrase from Dennis and Wartella (1996) as well as Pooley (2006).

school's elaboration that focused instead on social influence via interpersonal communication (Katz & Lazarsfeld, 1955). Despite some skepticism that such research should be considered within the domain of communication rather than sociology (Bauer, 1959), today interpersonal communication and social networks seem essential parts of the discipline even among those whose foci are elsewhere.

Nonetheless, I will argue that those raising alarms about minimal effects won the argument in at least one respect: They defined "effects," at least interesting ones, as those that involve categorical or directional *change* in attitudes or behaviors. When scholarly arguments about minimal effects resurfaced in the recent past (Bennett & Iyengar, 2008; Holbert et al., 2010), the locus of disagreement was on whether this type of media effect — in which attitudes or behaviors are made less extreme or shift into a new category entirely — was the only kind worth studying. Lang and Lang's (2006) reflection on the first minimal effects era stresses how, irrespective of the contested empirical basis for the *Personal Influence* findings, the effect was to narrowly define media effects as the kind the Columbia school deemed weak.

But all along, another kind of influence was acknowledged. Lazarsfeld et al. (1948), just after reporting the headline finding of "no overt effect," pose the rhetorical question of whether this means that campaign communications had no effects on the many people who voted along with their usual partisan identity. The answer:

> Not at all. For them, political communications served the important purpose of preserving prior decisions instead of initiating new decisions. It kept the partisans "in line" by reassuring them in their vote decision; it reduced defections from the ranks. It had the effect of reinforcing the original vote decision. (Lazarsfeld et al., 1948, p. 87)

And Klapper (1960), in his influential summary of media effects research, concluded, "[w]ithin a given audience exposed to particular communications, reinforcement, or at least constancy of opinion, is typically found to be the dominant effect" (p. 15). Lazarsfeld himself enumerated 16 distinct types of media effects (Lazarsfeld, 1948), which Katz and Lazarsfeld (1955) would say was not quite comprehensive. Lang and Lang (2006) later concluded that these many qualifications to the famous findings of minimal effects were not widely received or commented on by contemporary readers and collaborators, some of whom — like Berelson — would go on to declare the field largely played out. Many years later, Bennett and Iyengar (2008) would predict rather dismissively that in the present era, media are unlikely to "do anything other than reinforce prior predispositions" (p. 724), echoing the Lazarsfeld group (1948). This assertion prompted Holbert et al. (2010) to remind that "the study of persuasion involves analyses of response shaping, response reinforcing, and response changing processes of influence, not just the latter" (p. 17).

Perhaps the most prominent exception to my claim that stability is not treated as an outcome in communication research is inoculation theory (McGuire, 1961; Pfau & Burgoon, 1988). This approach comes from persuasion research and uses the biological metaphor of immunization to explain the way that persuasive messages can pre-empt competing claims by providing and then refuting counterarguments. The basic expectation, then, is that attitude or behavior change does not occur upon exposure to competing persuasion attempts when inoculation was part of the original message. Of course, since the point is to make successful attempts at persuasion robust to subsequent persuasive messages, this approach is still about change — just change that is more enduring. That being said, in applied settings, the focus is on

promoting the maintenance of pre-existing attitudes, in which case inoculation is not so different from my own approach beyond its focus on attitudes and mine on identity.

In political communication research, particularly the study of campaigns, there has been some interest in the apparent feebleness of attempts to persuade voters. For instance, election forecasters can predict the outcome of presidential elections long before the campaign begins (e.g., Abramowitz, 1988; J. E. Campbell, 1992; Lewis-Beck, 1985). If the campaigns were very influential, one would expect it to be more difficult to predict the outcome of the election based on the (mostly) economic indicators used in the forecasting models. Another aspect of this puzzle is figuring out why, if it is so straightforward to project the results in advance, do opinion polls typically fluctuate — sometimes dramatically — throughout the election season. Gelman and King (1993) argue the campaigns have the effect of supplying voters with the information they are ultimately expected to use to make an informed choice, like candidate partisanship, the economic conditions, and the effects of the policies of incumbents. This information is also typically part of the forecasts and therefore, in a way, the voters ultimately make decisions based on the same information the forecasts do. One could argue this is a case of the campaigns imposing stability on public opinion over time (see also Finkel, 1993). Holbrook (1996) makes a similar argument, describing the forecasted result as an "equilibrium" and arguing campaigns tend to push opinions toward the equilibrium level. In some respects, this is similar to the model I will propose although I will focus on the individual level of analysis and more general contexts that encompass non-election seasons. The equilibrium model, moreover, seeks to explain the *lack* of stability in a phenomenon of interest while I start by dealing with the apparent high stability of several constructs and build a model to explain it.

This dissertation will focus on an effect besides the narrow one that has defined the historical narrative of media effects research. It is the idea of reinforcement that most closely resembles the focus of this project, although in a subsequent section I will point to some inconsistencies in the apparent conceptualization of this term. To be more precise about my own aims, I will argue that *stability* of attitudes, identity, and behavior is an underappreciated potential impact of communication — both mediated and interpersonal. Although stability clearly has been a through-line of findings and debates in communication research, it is rarely been treated as an interesting end in its own right. The reasons for this include minimal theorizing about stability as an outcome of communication, the conflation of stability and polarization, as well as a lack of well-known methodological and statistical tools for the empirical study of stability.

Rather than specific attitudes or behaviors, I will focus on social identity as the construct whose stability is impacted by communication. The context for these identities will be politics, although the arguments, theory, and methods will largely generalize to any context in which social identities are relevant. The reason for the focus on politics, besides the general importance of understanding citizen behavior in democracies, is the increasing recognition of the centrality of social identity as cause of public opinion and political behavior in the United States (Huddy et al., 2015; Iyengar et al., 2012) and elsewhere (Brader & Tucker, 2012; Huddy, Bankert, & Davies, 2018). This research has shown that political behavior is much more likely to be explained by (partisan) social identity than specific attitudes or ideology, which themselves may be downstream effects of identification (e.g., Barber & Pope, 2019).

Stability vs. Reinforcement

Research and debate about persuasion has tended to contrast the outcome of conversion — in which someone starts out with one attitude or behavior and, because of some communication(s), adopts a new one — against reinforcement, which is understood as retaining the original attitude or behavior. The term reinforcement, however, has not been used consistently. Some use reinforcement to denote a lack of change or implying a resistance to change while others see reinforcement as a strengthening of the attitude or behavior. I use "strengthening" in this context to mean more extreme (e.g., a liberal political position becomes more liberal). Lazarsfeld et al. (1948) described the reinforcement effect as having "reduced defections from the ranks" (p. 87) in the context of voting for the candidate of one's political party. Voting is a binary behavior; a person either votes for the candidate or not. It cannot be said in this case that Lazarsfeld et al. (1948) showed that a person voted *more* for their party's candidate when exposed to campaign materials, only that in aggregate exposure was associated with greater numbers of people voting for their party's candidate. In other words, it is not clear whether reinforcement meant a strengthening of an underlying attitude or just that individuals were effectively inoculated against conversion. Both are possible, but a common theme in research is a lack of distinction between these possibilities.

Both Bennett and Iyengar (2008) and Holbert et al. (2010) agree conceptually that in political communication, an outcome of reinforcement would be polarization[4], meaning a reinforcement effect implies more extremity in the attitude or behavior. Dilliplane (2014) operationalizes reinforcement in the context of voting as an increase in favorability towards a candidate among those whose stated intention to vote for the candidate did not change throughout the political campaign. Livingstone (1996), on the other hand, comes out more clearly with a conceptualization of reinforcement as a lack of change, referring to such effects as those that "reinforce the status quo" (p. 307). Knobloch-Westerwick and Meng (2011) at times refer to both reinforcement and stability but do not make a clear distinction. In their discussion, they refer to reinforcement as the product of individuals seeking "self-consistency and stability" (p. 365). Klapper (1960) makes a distinction between reinforcement on one hand and "constancy" on the other (p. 15). Potter (2011), who was trying to bring clarity to the conceptualization of communication effects, introduced "weight" as a property of attitudes that may be affected by communication. In this framework, communication may result in categorical change (a change in kind), a change in strength, or weight, which Potter describes as its resistance to change and later refers to as reinforcement. Because this conceptualization does not relate clearly to any previous literature or psychological constructs, I will not return to the weight idea but merely introduce it as yet another example of the many related but distinct versions of reinforcement in communication research. Part of my operationalization of stability at the

[4] They do have an apparent disagreement over whether and to what extent such effects actually occur.

individual level, however, will be conceptually quite similar to Potter's idea of weight. In light of this terminological uncertainty, I will avoid "reinforcement" to describe the concept of interest here even if some will understand it as such. Instead, I will refer to this core concept as "stability," which more clearly communicates the phenomenon in which change of any kind is either small, temporary, or lacking entirely.

When I discuss stability, I will be referring to a lack of change, or consistency of identity, attitude, or behavior. In other words, strengthening as I have described it is not an example of stability, just as categorical changes are inherently instances of instability. Later, I will discuss in detail how to differentiate different types of change. I first will make a distinction a la Nesselroade (1991) between intraindividual development and intraindividual variability. **Development** is "more or less enduring" and "construed as developmental" while **variability** refers to "relatively short-term changes that are construed as more or less reversible and that occur more rapidly" than change as just defined (Nesselroade, 1991, p. 215). As I have described the theoretical-historical context of this study, most research interest is on change while I am focused on something more like variability as Nesselroade understands it. Note that stability, in this framework, is a purely within-person phenomenon. As I will review in the coming chapters, most research on the stability of communication focuses on rank-order stability. Rank-order stability refers to the extent a person measured at one time point is expected to rank similarly compared to others measured at the same time in a subsequent measurement. This will only correspond to the over-time consistency of the observed values under certain circumstances that are rarely considered. Such methods also discard non-monotonic variation as measurement error, even though it may correspond to true variation that just does not persist over time. In a self-

regulating system — something that should describe a person who is stable — a departure from equilibrium is not expected to last, much in the same way a measurement error is expected to behave. My approach to stability will involve looking skeptically at claims of large amounts of random measurement error and instead assuming that within-person variations may indeed reflect true variability.

Chapter 3. Social Identity

As the main construct under study in this dissertation and the most purely psychological, I devote the following pages to explicating social identity. I provide a brief intellectual lineage of the theory behind social identity research, which in the beginning occurred in the context of laboratory experiments using contrived groups. Although this is quite different than the political identities I will study, this research provides much of the early evidence base and demonstrates a key point meant to generalize to real-world identities: Intergroup conflict is not solely due to competing rational interests, but instead is due at least in part to a psychological predisposition towards group-based thinking. I then move on to break the social identity construct into constituent parts: the identity itself (categorical), the strength of the identity, as well as identity threat and related concepts such as collective self-esteem.

Social identity is a term that is almost self-defining, but in social psychology and now many other social sciences, it implies a specific intellectual lineage. The bulk of academic research on social identity is rooted in Social Identity Theory (SIT; Tajfel & Turner, 1979) and its extension, Self-Categorization Theory (SCT; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). The two theories are often collectively referred to as the "social identity approach" (e.g., Hornsey, 2008, p. 205) and in absence of further clarification that is what I will be referring to when I discuss social identity here. Put succinctly, social identity is the "part of an individual's self-concept which derives from his knowledge of his membership of a social group (or groups) together with the emotional significance attached to that membership" (Tajfel, 1974, p. 69). This is not the only type of identity one may have. For instance, one may hold personal identities; considering oneself a friendly person does not imply group membership (even if there are many

other friendly people) but rather an individuating characteristic (Brewer, 1991). One could plausibly think of him/herself as a member of a group of friendly people, in which case the social identity principles apply (i.e., group dynamics may override the objective basis for the group, like deciding to be rude to unfriendly people to advance group interests). The legitimacy of a social identity depends only on whether it is psychologically real, felt by the individual to represent a relevant group to which the individual feels they belong. This latter point is part of the theoretical developments stemming from SCT, which elaborated the cognitive processes underpinning SIT, focused moreso on the conceptualization of social identity and the behavioral manifestations thereof. This is also consistent with the political science approach to partisanship, whose foundational text notes "[t]he active membership group that forms the core of each party is small; the psychological group, however, is very large" (A. Campbell, Converse, Miller, & Stokes, 1960/1980, p. 295). Another useful aspect of focusing on individual perceptions of group membership is avoiding the need to impose standards on what constitutes membership. Although early SIT research assumed that self-stereotyping — attributing traits typical of group members to the self to feel more a part of the group — was a necessary part of identification, subsequent research has found that self-anchoring — attributing traits of the self to other group members — is another way people come to identify more closely with a group (van Veelen, Otten, & Hansen, 2011).

I will refer repeatedly, as Tajfel (1974) did, to the self-concept. This is a fundamental construct in social psychology that refers to the way a person views him or herself. A simple way to think of the self-concept is that it is the answer to the question, "who am I?" In fact, early explication of the construct built on the many answers to this question (Kuhn, 1960). Even prior

to theories of social identity, it had been established that a major aspect of how people think of themselves is as members of social groups (Kuhn, 1960; Rogers, 1959). This understanding of the self in terms of group membership is not inconsistent with the symbolic interactionist conceptualization of the self (e.g., Mead, 1934). As is central to the symbolic interactionist view of self, social identities are part of the self that is formed through social interaction and help people situate themselves in the society (see also Festinger, 1954). Compared to symbolic interactionism, the social psychological tradition has placed more emphasis on the implications of the self-concept for self-esteem. Most psychological theory assumes that people are motivated to have a positive self-concept, although this motivation can be in tension with objective reality. Inability to reconcile reality with a positive self-concept is expected to lead to low self-esteem; distorted views of self or reality may correspondingly lead to pathologically low or high self-esteem. I will not be studying the self-concept per se, but social identities are a part of the self-concept. What I will not deal with are the other descriptive aspects of the self-concept, like the attributes one ascribes to oneself. Psychologically, interactions are expected to vary in the extent to which one acts "in terms of self" (interpersonal interactions) or acts "in terms of group" (intergroup interactions; Turner & Reynolds, 2012, p. 403). At the interpersonal end of the spectrum, an individual perceives his or her discussion partner only in terms of their personal characteristics (e.g., that person is friendly, confident, etc.). At the intergroup end, the perceiver sees the interactant as a stand-in for what is known about the social group to which the other person appears to belong. In practice, such extremes are probably rare but the creators of the approach expected that intergroup perceptions likely predominated. When individuals are engaging in intergroup thinking, they not only perceive others but also themselves in terms of

their group status. The salient aspects of one's self-concept will depend on the extent to which social or personal identities are activated in the situation. And despite the nomenclature I have used here, the principles of intergroup interactions can also apply when the people involved see one another as members of the same group (i.e., intragroup interactions).

There are several substantive consequences of intergroup thinking, by which I mean thinking of self and others in terms of group membership. Those perceived to belong to groups other than one's own are depersonalized; that is, they are seen as more or less interchangeable with other members of the group. Likewise, individuals may engage in self-stereotyping, acquiescing to the norms of the salient group to which one belongs. An important contribution of the social identity approach was demonstrating the persistence of group-oriented behaviors even when the objective basis for the categories was very thin. Studies demonstrating this are part of the so-called minimal group paradigm and involve placing experimental participants into arbitrary groups, such as "overestimators" and "underestimators" following a task in which they are asked to estimate the count of a large number of objects in a short time (e.g., Tajfel, 1970). In follow-up tasks, participants engage in what amount to economic games and the pattern of results are what is often described as irrational; participants make choices that maximize gains relative to the (arbitrary) outgroup rather than maximize absolute gains. For example, if a participant is told they are a part of the group of overestimators, they show a tendency to reward fellow overestimators — who are unseen by the participant and are a fiction of the experimental setting — but will reduce the total reward to in-group members if it makes punishments to out-group members, and therefore relative gains for the in-group, possible. Experiments in this tradition, which comprise much of early social identity research, might seem overly removed

from reality. They built the foundation for this psychological approach to intergroup behavior, however, because they provide evidence that these behaviors are not necessarily the product of real competition but are instead a basic psychological process that persists even when the groups are artificial and the stakes are low/non-existent. An implication of these minimal group findings is that the conflict between real-world groups cannot be eliminated by simply addressing whatever material conditions that may have brought the two into conflict in the first place. In politics, this means that negative partisanship is unlikely to be ameliorated by having the parties share interests. One could argue the two major U.S. parties have many shared interests but often fail to act on them because of their resistance to cooperating with outgroups.

A tradeoff of the minimal group paradigm's focus on the internal validity claims made possible by the laboratory setting and artificial groups — results can hardly be explained by pre-existing characteristics of the participants or unappreciated aspects of the history of the groups — is some of the nuance and natural variation is lost in the process. For instance, for real-world groups, it is important not just to consider that one identifies with the group, but also how strong the identification is. In the minimal group setting, identification can be assumed from the experimental induction and meaningful variation in strength is not expected because of the contrived groups. I conceptualize strength of identification to be primarily about belongingness, in which a person feels they are clearly a member of the group and their membership in the group is an important part of the self-concept. Although it is expected that the strongly identified feel positively towards the group, this may not always be true as will be discussed in the next section.

The need for consideration of identity strength is better understood in the political science research on partisanship, which precedes SIT. In *The American Voter*, the authors write "[m]ost Americans have [a] sense of attachment with one party or the other. And for the individual who does, the strength and direction of party identification are facts of central importance in accounting for attitude and behavior" (A. Campbell et al., 1960/1980, p. 121). Their measure that asks partisans a binary follow-up question about whether they are "strong" or "not strong" partisans is by far the most used measurement of identity strength in political science today. Psychologists have since developed scale measures of identity strength, designed for any social identity (Mael & Tetrick, 1992). These ask respondents to agree or disagree with items like, "when someone criticizes [the group], it feels like a personal insult." Such measures have since been adapted to partisan identity in the United States (Greene, 2002; Huddy et al., 2015) and other countries, including multi-party systems (Bankert, Huddy, & Rosema, 2017). The categorical dimension of social identity has been assumed trivial to measure — a claim I will not dispute — and has a tradition in multiple fields of being measured with simple yes/no or checklist questionnaire measures.

The way party identification is most frequently measured in American politics research putatively taps into strength, but has some important drawbacks. Respondents are first asked whether they are Republican, Democrat, Independent, or something else (slight variations on the response choices after Republican or Democrat are not uncommon). Those who choose Republican or Democrat are asked in a follow-up if they are a "strong" or "not strong" Republican/Democrat. Those who did not choose Republican or Democrat are given a follow-up asking if they "lean towards" or are "closer to" one or the other party, with the option to indicate

they are not closer to one or the other. The responses to these items are typically analyzed as a 7-point scale ranging from strong Republican, not strong Republican, lean Republican, no lean, lean Democrat, not strong Democrat, to strong Democrat (whether Republican or Democrat comes first is unimportant and various labels are used for the middle point). This has face validity but in practice has revealed problems. In particular, one would expect that those who do not choose Republican or Democrat in the first question are not identified with a party. Instead, research has shown the "leaners" at minimum act quite a bit like partisans (J. Dennis, 1992; Klar, 2014; Petrocik, 1974). Leaners have been shown to be as voracious of media consumers as strong partisans with media diets just as biased to their preferred party (Long, Eveland, & Slater, 2019). Petrocik (2009) puts it bluntly: "Leaners are partisans" (p. 562). He explains that this tendency to first choose independent identification before admitting a preference, predominantly done by middle-class whites, is selective self-presentation to emphasize the ability to think for oneself rather than take cues from party leaders. In terms of political behaviors and attitudes, leaners are at least as partisan as "not strong" partisans and perhaps moreso. Theodoridis (2017) finds, using implicit measures, that leaners are clearly distinct from independents, Democratic leaners show about as much preference as "not strong" Democrats, and Republican leaners show more party preference than "not strong" Republicans. It appears "independent" is often a statement of a personal identity that does not preclude holding the partisan social identity as well. There may also be cases in which a person identifies with an allied or similar party in addition to the major party (e.g., some people choose "other" and then explain in an open-ended response that they identify with the Tea Party, which is not formally a political party, or Libertarians,

which is but is more often part of Republican coalitions than otherwise). I take the position that we ought not come to strong conclusions about strength of partisanship based on this measure.

Identity Threat and Collective Self-Esteem

Identity threat is an important part of the theoretical approach I will introduce later. Partly due to the experimental tradition in which the identity threat construct is manipulated by the researcher, there is relatively little work on the direct measurement of identity threat. Researchers of identity threat have generally taken the position that people cannot be reliable reporters of their experiences of identity threat both because they may not recognize threats as they occur and due to the threats being embarrassing to admit when they are recognized (Scheepers & Ellemers, 2005). Taking cues from the experimental approach, field researchers can instead try to identify threats in the social environment and look for indications that they have occurred or make comparisons to people presumed not have been under threat. I am introducing the collective self-esteem construct as such an indicator of identity threat. I will also briefly discuss another, related construct known as collective narcissism.

As discussed earlier, part of the motivational basis for maintaining one's self-concept is the maintenance of self-esteem. A contribution of the social identity approach has been making the distinction between the collective and individual parts of the self-concept, both of which have ramifications for self-esteem. When circumstances cause group memberships to reflect negatively on the self, one may take solace in — and emphasize — the personal parts of the self-concept that are not threatened. This is not to say there is always a clean separation of personal

and group parts of the self-concept; for instance, a political identity may be intimately tied to one's sense of being a moral person. It is still useful to make a distinction between the social group-oriented aspect of self-esteem from the personal, or overall, levels of self-esteem. Conceptually, I am just introducing a name for something that is fundamental to the social identity approach. Collective self-esteem, in layman's terms, is how positively or negatively one feels about the groups to which they belong. Although there is some variance in conceptualization of this construct, for my purposes it is about the extent to which a person feels the group is a good one to belong to. This is partly based on perceptions that others view the group positively as well as actual events like (in the case of politics) elections, scandals, and achievements. Operationally speaking, it is important to be able to distinguish between the overall evaluation of the self-concept (i.e., self-esteem) and the part concerning social identity in particular (Crocker & Luhtanen, 1990).

Identity threat did not emerge as a distinct, named construct in early research from the social identity approach but was the motivational basis for many of the theoretical predictions. SIT drew on Festinger's (1954) Social Comparison Theory, which suggested that group consciousness is almost fundamentally comparative. The purpose of holding a social identity, in the social identity approach, is both for sense-making (where do I stand in the world?) and for self-enhancement. The social groups to which one belongs should make them feel more positively about their self-concept. In that sense, then, comparisons between groups should reflect positively on the in-group. The original incarnation of SIT provides the framework for conceptualizing identity threat and its consequences. Tajfel and Turner (1979) explained that the connection between status differentials and intergroup competition is neither direct nor

consistent; they argue the connection is mediated by "social identity processes" (p. 43) laid out by SIT.

Assuming people are motivated to maintain their self-esteem, which is inextricably bound to salient social identities, the original incarnation of SIT suggests several possibilities when people are confronted with an upward social comparison (their salient social identity group is inferior to another salient group). If social mobility is possible, people may simply try to join the superior group. The fact people do not seem to change their identity despite ample opportunity and apparent motivation required more explanation, however. Instead, the lower-status group member may intentionally try to temporarily reduce the salience of that social identity and gravitate towards other social identities (trying to improve collective self-esteem) or to personal identities (emphasizing tying overall self-esteem to self-esteem at the personal level). Alternately, because strength of identification is based so much on the extent to which the group identification is compatible with the self-concept, one may ultimately reduce strength of identification, thereby restoring collective self-esteem. Over time, a person may begin to see himself or herself less as a member of the group because they cannot find a way to integrate that membership into a positive self-concept. My argument, then, is that collective self-esteem and identity strength can be in tension with one another; a person will struggle to simultaneously maintain high identification and low collective self-esteem. Ultimately, collective self-esteem will need to be managed by either reducing identification with the threatened group or reducing the threat.

An example from the study of sports fandom, which has since incorporated SIT, is the predicted reactions to team success and failure. When a favored team is victorious, fans are

known to engage in what is called "basking in reflected glory" (BIRGing; Cialdini et al., 1976). A simple manifestation of this is in language use: fans of the winner will use phrases like "we won" which is a fairly literal way to express membership in the target group. The opposite phenomenon came to be known as "cutting off reflected failure" (CORFing; Snyder, Lassegard, & Ford, 1986). The corresponding language example for CORFing is "they lost," in this case expressing distance from the target group. These basic findings have been generalized to politics, showing results consistent with BIRGing and CORFing in tweets after the Scottish referendum (Lachlan & Levy, 2016), candidate ratings in the USA (Miller, 2009), and post-election display of political yard signs in multiple countries (Boen & Vanbeselaere, 2002; Boen et al., 2002; Miller, 2009).

Another strategy when confronted with negative social comparisons is what Tajfel and Turner (1979) call "social creativity" (p. 43). This, in layman's terms, corresponds to "moving the goalposts." That is, the basis of the comparison between groups is changed such that it is no longer threatening. In politics, this may mean focusing on litmus test types of issues — a pro-life Republican whose party lost an election may rationalize their loyalty (quite reasonably) by pointing to how incompatible the opposition's position is with theirs on something so important. In other words, the focus shifts from political standing to moral standing. This is quite plausibly an important role of communication, perhaps moreso for media than interpersonal discussion. Partisan outlets, in particular, will provide audiences with these sorts of rationales for continued identification, through some mixture of examples of the in-group's strengths and out-group's problems. Assuaging co-partisans is likely a strategy pursued by elites via non-partisan media as well. Yet another strategy is to pursue direct competition when such competition is possible.

Sports fans, as an example, always have future seasons to look forward to in which the standings may be re-ordered in a way favorable to one's preferred team. Regular elections in the political sphere give discouraged partisans reasons to believe the existing state of affairs may be changed next time around. A lost election may also lead to claims that the contest itself was unfair. When George W. Bush was declared the winner of the 2000 US Presidential election after bitter court battles over the counting of ballots in Florida, Democrats predictably reacted in part by claiming the election was stolen from them. Although this does not change the outcome, collective self-esteem is preserved by interpreting the situation as one that showed Gore *should* have won. Another strategy in the face of political defeat is to scapegoat the losing candidates or others deemed responsible for the loss, focusing on strategic decisions or deviations from group norms as the problem. The overarching theme is that people pursue strategies that preserve self-esteem, either by trying to (re)cast the salient social identity in the positive or by reducing the strength of the problematic identity.

Luhtanen and Crocker (1992) created an oft-used scale measure of collective self-esteem. The scale originally was used to evaluate all groups one was a member of, but it has been almost always reworded to address particular groups of interest (Crocker, Luhtanen, Blaine, & Broadnax, 1994; Rubin & Hewstone, 1998). The scale as originally proposed has 4 dimensions the authors call *membership*, *public*, *private*, and *identity*. My focus is on *public* and *private*. *Private* regards the individual's affect towards being a member of the group (example reverse-coded item: "I often regret that I belong to [the group]") while *public* taps into assessments of the group's status (example item: "overall, [the group] is considered good by others"). *Membership*, not of great interest for the present study, refers to feelings of being accepted in the group.

Identity is actually a measure of strength of identity, which can be useful in assessing how much a particular identity is likely to impact global self-esteem but is clearly not itself an aspect of collective self-esteem. They report sometimes significant but relatively modest correlations of the *identity* subscale with *public* (from .05 to .30) and somewhat stronger correlations with *private* (.41 to .53) across three studies. Some research has used the *identity* subscale as a measure of identity strength (e.g., Abrams & Giles, 2007; Luong & Knobloch-Westerwick, 2017).

Regardless, I want to emphasize that strength and collective self-esteem are distinct (albeit related) constructs and despite an expectation they be correlated — it only makes sense to identify more strongly with groups that are judged as good — they can vary independently, especially in the face of identity threat. This point has been made elsewhere as researchers have attempted to divine the aspects of the scale that tap into collective self-esteem rather than related constructs (Bergami & Bagozzi, 2000; Ellemers, Kortekaas, & Ouwerkerk, 1999). There is some disagreement in whether *private* (Cameron, 2004; Ellemers et al., 1999), *public* (Hunter et al., 2005), or both (Sellers, Smith, Shelton, Rowley, & Chavous, 1998) best represent the construct. I see them as too closely related to measure just one, as a person may be able to recognize low status (*public*) without feeling badly about being a group member (*private*) and may experience shifts in *private* that are unrelated to the perceived status of the group. Unfortunately, there is not sufficient research and theory-building on these dimensions of collective self-esteem to have precise predictions about their causal relationship, if there is any. Sellers et al. (1998), thinking about the context of the racial identity of Black Americans, suggest that people may be able simultaneously have high *private* regard for the group while simultaneously understanding the

group is not held in high regard by others (*public*), although they do not have data on the time dynamics of these two subdimensions. There is, however, no denying that many Black Americans feel very positively about identifying as Black but also perceive the group's status to be low in the eyes of others. What is not so clear is whether a drop in *public* from whatever baseline level would coincide with a drop in *private*, from whatever its baseline is. I treat them both as parts of the overall collective self-esteem construct.

Chapter 4. Communication and Identity

My approach to the link between communication and identity will focus on a configuration that is not often studied: treating communication as a cause of identity, particularly communication affecting the strength of identity and collective self-esteem. In this chapter, I will explore theoretical mechanisms for communication to affect identity and review some of the extant research on communication and identity. I am dealing with social identity at the individual level of analysis, so I will not study in any detail the role of communication in creating group norms, altering public perceptions of the group, and so on. This means some of the most prominent approaches to communication and social identity, like the Communication Theory of Identity (Hecht, 1993; Jung & Hecht, 2004), are not necessarily contradicted by the claims I am advancing but not particularly relevant for the research questions. Similarly, the Social Identity Model of Deindividuating Effects (SIDE; e.g., Postmes, Spears, & Lea, 1998) is focused on how identity, interacting with qualities of a medium, affect the content of inter/intragroup communications. In the same way, Communication Accommodation Theory (Giles, Coupland, & Coupland, 1991) is quite invested in identity but mostly on how identity affects the content of (interpersonal) communication and the relational consequences thereof.

Other relevant approaches in communication research have tended to focus on attitudes but the findings in many cases have clear applicability to identity and in some cases may have unknowingly reflected identity processes all along. Research on motivated reasoning and related processes theorize attitudes to be a moderator of communication effects (Kunda, 1990; Taber & Lodge, 2006). The same is true of hostile media effect research (Giner-Sorolla & Chaiken, 1994;

Hastorf & Cantril, 1954; Reid, 2012). Theories of selective exposure, again more oriented towards attitudes than identities, generally position the attitude or identity as cause rather than effect (e.g., Knobloch-Westerwick & Meng, 2009). The selective exposure approach does differ in that it has developed to consider both that identity drives selection of communication (primarily media in this line of research) but also that as a consequence of selection, the underlying identity or attitudes may be strengthened, thereby increasing the likelihood of subsequent selection (Knobloch-Westerwick & Meng, 2011; Stroud, 2010). Another useful aspect of the selective exposure approach is the importance it places on the distinction between types of communication content. Theories of selective exposure presume a preference for communication content that is consistent with pre-existing attitudes. Relatively recent refinements, at least in politics, demonstrate a clear preference for media sources that favor one's own party (Iyengar & Hahn, 2009; Stroud, 2010) but also that people may not explicitly avoid disagreeable sources (Frey, 1986; Garrett, Carnahan, & Lynch, 2013).

The original explanation in the selective exposure approach for why people prefer attitude-consistent communication came from cognitive dissonance theory (Festinger, 1957), which argues it is cognitively uncomfortable to be confronted with information contradicting one's own views. This mechanism better explains avoidance of disagreeable information, which is not empirically supported and may only occur under certain conditions (Druckman, 2012; Valentino, Banks, Hutchings, & Davis, 2009). The motivational basis for selective exposure is not clear given the lack of selective avoidance, since there must be more reason to choose congenial communications than mere avoidance of dissonance. One such explanation may be the aforementioned motivated reasoning literature, which originated partly from research on

confirmation bias. A motivation for congenial information in this case would be based on a desire to be reassured of one's correctness. Unfortunately, this approach by and large has had an attitudinal (or ideological) orientation, such that there is a focus on learning and attitude-specific discrepancies. Political research has had a tendency to conflate ideology and partisanship, due in part to the two-party system in the United States that conforms well to liberal-conservative distinction as well as an assumption that people become partisans because of their ideology (and when their ideologies mismatch their partisanship, it is because they are not knowledgeable; Converse, 1964). As mentioned previously, there is evidence that the opposite is true (Barber & Pope, 2019) and that in any case political behavior is often better explained by partisanship than issue attitudes or ideology (Huddy et al., 2015). The main point is that although there is much to be learned from the selective exposure paradigm, its theorized mechanisms do not clearly show why social identity would drive media selection besides members of identity groups holding a set of attitudes they consider relevant to belonging to the group.

A more compelling explanation for why communication choices may be driven by social identity is that certain forms of communication may promote positive feelings about the group and/or membership in it. Blumler (1985) extended uses and gratifications theory (U&G; Katz, Blumler, & Gurevitch, 1973) to enumerate a social identity gratification of media use, later tested by Harwood (1997, 1999). In other words, media may "provide positive social comparisons with outgroups" (Harwood, 1999, p. 125). This could be, as Harwood (1999) conceptualized, by portraying group members positively or perhaps just as or more often in politics, the explicit contrasting of in-group and out-group in a way that is favorable for the media consumer. This account does not exclude the attitudinal explanations provided by

selective exposure theory, since particularly in politics there are values and issue attitudes that are considered important aspects of group membership. At any rate, the expectation is that these portrayals can increase strength of identification and collective self-esteem[5].

Abrams, Eveland, and Giles (2003) build on Harwood (1999) and suggest identity-gratifying media could promote subjective group vitality. To explain the significance of this assertion, I will need to digress to introduce this concept. Vitality theory (Giles & Johnson, 1987) originated in an interest in language use (and non-use) among minority and threatened ethnic groups. Vitality referred to the objective status of the group, in terms of political power, demographics, or cultural capital. Subsequent developments in the theory came to recognize and emphasize what is referred to as subjective group vitality, which refers to individual assessments of group status as just defined. Although once conceptualized differently, vitality researchers have recently come to believe that subjective vitality — at least as it is commonly measured — is more of an evaluation of the group than it is a measure of a person's knowledge about objective status (Abrams, Barker, & Giles, 2009; Smith, Ehala, & Giles, 2017). Smith and colleagues (2017) say subjective group vitality is tantamount to the group comparisons made in social identity theory. In this way, I would argue that one should think of subjective vitality, when

[5] Harwood (1999) does not find a "significant" correlation between collective self-esteem and retrospective reports of identity-consistent media but does not report the actual observed values in the relatively small student sample used. The identity category used in the study, age, may not map particularly cleanly onto politics and other social categories either.

applied to a person's own social groups, as quite similar conceptually to collective self-esteem. Returning to the Abrams et al. (2003) assertions, this is perhaps the first claim in the communication literature that media may promote collective self-esteem, at least insofar as collective self-esteem and subjective vitality are closely related. Reid, Giles, and Abrams (2004) subsequently developed the Social Identity Model of Media Usage and Effects, which formalizes these predictions about media use and subjective vitality, not all of which are relevant for the questions at hand. This theory regretfully has not quite taken off in terms of inspiring new research, as measured by citation impact, outside of work on vitality theory. Nevertheless, at least in broad strokes this approach is consistent with the model of communication effects I will be advancing in this dissertation.

Another general set of findings it is important to address concerns the extent to which partisans avoid communication that is not identity-affirming. Particularly since the internet became a greater part of daily life, concerns have been raised about the potential for people to completely isolate themselves from dissenting views (e.g., Sunstein, 2001, 2007). As discussed in the context of selective exposure, however, these predictions have not been well-founded in empirical research on media (Bakshy, Messing, & Adamic, 2015; Dubois & Blank, 2018; Flaxman, Goel, & Rao, 2016; Gentzkow & Shapiro, 2011). There is a largely separate literature on interpersonal discussion that tends to concur that people indeed tend to have diversity in their political discussion networks (Eveland et al., 2018; Huckfeldt et al., 2004). Comparatively, Gentzkow and Shapiro (2011) document greater homogeneity in social networks than media diets, although both were considered diverse. This raises questions about the role of identity-affirming communication in the context of an overall set of communication behaviors that, on

average, are not wholly affirming. Research on selective exposure would suggest people may simply have different motivations for communication that is not particularly affirming or even outright hostile. Regardless of the initial motivation for exposure, motivated reasoning research shows people to be willing to expend more effort counterarguing with disagreeable information. For these reasons, my focus is on the amount of identity-affirming communication, irrespective of other identity-relevant communication. Likewise, in the case of social network data, I treat the partisan composition of the network to be an approximate surrogate of identity-affirming communication. I do not have strong *a priori* expectations that the percentage of a social network that has the same partisanship as the person in question is particularly important when it comes to identity-relevant communication. A person with a large, diverse network will engage in more identity-affirming communication than someone with a very small, homogenous one. It is possible that diverse networks could be threatening, but it seems clearer that the broader social environment has this effect than the more proximal discussion network (Long et al., 2019).

Reinforcing Spirals Model

The Reinforcing Spirals Model (RSM; Slater, 2007, 2015) is useful for integrating these ideas. Put briefly, the RSM argues for treating communication and constructs often treated as outcomes of communication (like attitudes and identities) as endogenous parts of a system. It serves to integrate media effects theories (in which communication is the independent variable and attitudinal variables like identity are outcomes) and selective exposure theories (in which attitudinal variables like identity are independent variables and communication is an outcome) by treating communication as both cause and effect. RSM further argues for taking cues from

systems theory to manage that integration. Speaking speculatively, much of the heuristic appeal of the RSM to the average communication researcher is the explanation it provides for how some people become extreme in both their communication habits and attitudes. In systems language, this results from positive feedback loops in which communication causes more extreme attitudes, identities, and/or behaviors, which in turn cause more selective or frequent communication in the same domain. As an example, among the studies that inspired the formalization of this theory was one in which violent media exposure among adolescents appeared to increase aggressiveness while increases in aggressiveness also appeared to increase violent media usage (Slater, Henry, Swaim, & Anderson, 2003). Such a pattern of results suggests the possibility of the positive feedback loops that characterize some of the theory's heuristic appeal. Empirically, studies that claim to be testing or implementing RSM are often panel designs in which communication (usually media exposure) and some attitude or affect both have positive coefficients for their effect on the other in a cross-lagged regression model. Taken very literally, such results imply ever-increasing extremity in both communication behavior and attitude.

I will focus on another aspect of the RSM. Again taking on the language of the systems theory, Slater (2007) states that the norm is for the system to be self-regulating, rather than purely mutually reinforcing. In other words, usually people's identities and attitudes along with domain-related communication do not become progressively more extreme over time. Instead, and as I have shown in the preceding pages, these are constructs that are quite stable. People, according to this logic, typically maintain things as they are. In fact, identity-affirming communication is described as "maintenance" of the identity in the explication of the RSM, a term that captures the expected result: stability. The systems explanation for why this is the case

is that our social systems are usually open. In other words, a person is exposed to more than just ideas and activities that push towards existing identities and attitudes. One is also exposed to counterinfluences that call those identities and attitudes into question. Moreover, people are multi-faceted and experience the pull of other interests, identities, and so on. Once enough affirmation for one identity is achieved, rather than pursue it further and become more extreme, the norm is to go onto something else.

Take for example a person who identifies both as Republican and a running enthusiast. The desire to run, learn more about running, and talk with other runners are things that occupy time that might otherwise be used to intensify the Republican identity. Beyond the time and cognitive constraints of having multiple interests, the kinds of ideas one may be exposed to when spending time with fellow runners may be inconsistent with the group values of Republicans, which could serve to moderate the Republican identity as a response to the identity threat of countervailing information. The micro-level view of the process is basically that one's Republican identity gets a bit stronger after watching *Hannity*, but recedes a bit in the intervening time due to discussions with others or other factors before the next episode (or other pro-Republican communication) brings the identity strength back up to normal. As in the case with a negative feedback loop, there are countervailing processes that effectively cancel each other out once equilibrium is reached. This implies there are indeed effects of media exposure and interpersonal conversations, but they tend to occur in a context in which the person uses them to avoid change. It would take a confluence of unusual factors to make increasingly extreme identity and communication levels the likely outcome.

This claim that these constructs tend to exist in a self-regulating system is an appealing one for several reasons. First, it applies the same type of logic to the processes that prevent change as it does for the processes that cause change. Second, it squares the theory with reality; most people, most of the time, in most domains do not have extreme attitudes, display extreme pro-group behavior, and do not engage in highly selective communication. Third, it explicitly theorizes about how and why key concepts will *not* change in a way that suggests certain types of empirical tests. This last point is important because I could come up with a theory that things do not change and, to test it, do run-of-the-mill media effects style statistical tests and claim null results as confirmation of the theory. My goal is to advance our ability to do theory testing in the domain of social identity maintenance, in which these forces act in concert to cause stability. The defect in many other theories, in my view, is the incuriosity about stability and whether it is in and of itself dependent on something or just a default state.

Some research on the RSM — and that influenced the RSM — has focused on volatile parts of the lifespan, like adolescent aggression (Slater et al., 2003), smoking (Slater & Hayes, 2010), and political interest (Moeller, Shehata, & Kruikemeier, 2018). These are times when the positive feedback loops are most likely to be observed since there is inherent instability in these constructs at this part of the lifespan (e.g., Jennings & Markus, 1984). In this way, we can see the reinforcing spirals as a mechanism for political (or other kinds of) socialization. Most of the time, except when populations are selected specifically for their life stage or other circumstance that is expected to be particularly volatile, people who will be studied have presumably already reached a relative equilibrium. Although there may always be some people subject to the positive feedback processes, a typical adult will be in a state of relative stasis. This could be why, for

instance, a study trying to connect local news use and community attachment failed to find evidence of a causal relationship between the two despite a meaningful cross-sectional correlation (L. H. Hoffman & Eveland, 2010). It may be the case that adults who are well-established in a community have reached that equilibrium and to detect the expected relationship, the sample would need to target people who have recently moved. This dissertation will not focus on the socialization processes that bring people to the dynamic equilibrium they have presumably reached in adulthood.

In general, the evidence about these key concepts suggest a certain type of relationship. Given the stability of media use and social networks, which I will review in more detail in the forthcoming chapter, the extent to which they are usefully conceptualized as primarily dependent variables is in doubt (granting that these variables are surely multiply and mutually caused to some degree). Categorically, social identities do not vary a great deal either. There are sufficiently many other explanations that the mere fact of stability in the key variables is not in and of itself evidence for the RSM, but it is at least consistent with the notion that social identities, social networks, and media use are co-equal actors in a self-regulating system. The question, then, is where is the action? There are several constructs subsumed under the umbrella of social identity that may capture hidden variation lurking beneath the categorical stability of social identity.

When collective self-esteem suffers, a person can go along one of a few paths, completely consistent with social identity theory. The road least traveled, per the available data, is to change allegiances to a new social group. One can also engage in identity-affirming communication, such as selective exposure to media sources in which the identity is validated. Examples could

include a Republican watching programs on Fox News, where the accomplishments of Republicans are frequently touted and failures of Democrats heavily scrutinized. A person could turn to a fellow group member to talk as well, in a way that is consistent with research on how social networks provide social support. The main remaining alternative, besides changing groups or affirming the identity via communication or other group-related behaviors, is to reduce the strength of identification with the group to alleviate the threat, similar to the predictions of cognitive dissonance theory (Festinger, 1957). This is basically a less extreme version of dropping the identity, but will not always be captured by research designs focusing on categorical change.

Another theory, not inconsistent with the RSM in this regard, suggests a motivational basis for consistency. Self-verification theory (Swann, 1983) asserts that not only do people have a desire for positive self-concept — the self-enhancement motive for responses to identity threat in the social identity approach — people are also motivated to maintain the self-concept even when self-esteem is reduced because it is so uncomfortable to change the way we see ourselves. I raise this because a key finding of self-verification theory is the use of social interaction to reaffirm the self-concept when one is confronted with information calling into question its accuracy. It also suggests an explanation for why it is the case that people do not simply change groups when their social group loses status (when switching is technically possible, as in the case of political parties). Although there is good reason to believe there is a self-enhancement motive in responses to identity threat, the self-verification motivation may explain why identities remain so "sticky" when they sometimes do not seem to confer positive self-esteem.

Decay

The RSM provides an elegant explanation for why extremity is not the norm: People have multiple interests and identities and even when those may be more or less aligned, they still live in a social context rife with moderating forces that exist in a diverse society. But as may have been apparent in my example of the Republican running enthusiast, it can become difficult to enumerate even in a hypothetical precisely what will cause one's identity to moderate in between identity-relevant communications. If a person never engages with politics except watching MSNBC's Joy Reid every weekend, it would be hard to say within the RSM what happens during the week that is not related to politics that would stop the Democratic identity — and the appetite for more pro-Democrat communication — from growing more extreme with every airing of Reid's program. And yet if I stipulated that a person identified with Democrats ignored politics except for a once-weekly partisan TV show, I doubt many scholars of political communication would expect such a person to be on a clear path to highly-selective partisan selective exposure and a very strong identity because the baseline level of communication seems too low.

I assert that, even in the absence of threat, social identities may have a "use it or lose it" quality. In other words, there is an inherent need to engage in identity-affirming activities, like communication, to maintain the strength of identities. I am not aware of this specific claim having been advanced or tested in the psychological literature, but it has face validity and can provide a simpler (or simplified) explanation to why and how communication and identities ultimately self-regulate rather than spiral under normal circumstances. The idea is that it is an inherent quality of identity that it needs active maintenance rather than maintenance only being

required because of persistent threats brought on by external forces or competing identities. Without any identity-affirming activities, it is hard to believe a person could continue to hold a strong attachment. In my (admittedly arbitrary) example above, the reason one does not read this hypothetical person as at risk for a positive feedback loop is because a person who engages with politics so infrequently is very unlikely to perpetually increase the strength of their identification even if there are no obvious threats to the identity to confront. I will refer to this temporal aspect as decay, or a basic tendency for identity strength to trend towards zero absent any identity-affirming behavior.

This is consistent with Slater's (2015) invocation of chronic accessibility (Fazio, Powell, & Williams, 1989) as an outcome of media selection. I am suggesting it also may be a cause insofar as the attitude accessibility tradition shows accessibility can sometimes *cause* attitude strength (Roese & Olson, 1994), especially when closely linked with identity (Boninger, Krosnick, & Berent, 1995). Another cognitive explanation for why this might occur and why it would lead to communication comes from self-verification theory. A person engaging in self-perception (Bem, 1972) may question the accuracy of their self-concept if that self-concept includes a social identity that plays an insufficient role in daily life. Identity-affirming communication, then, can bring the self-concept back in line with reality. In the self-verification perspective, identity-affirming communication would usually be preferable to changing the self-concept.

Identity maintenance is needed, in my view, to counterbalance decay and keep the identity as part of the self-concept. How much is needed? I would argue that this is related to the strength of identity. By way of analogy, consider the physical law that an object cools faster

when it is much hotter than the ambient environment. The rate of decay for social identity may be similar: the stronger the identity, the more identity maintenance is needed to counterbalance the inevitable decay. People reach an equilibrium in which their media use and social contacts are just identity-consistent enough to counterbalance the decay. This can make for an easier explanation of why people, obviously limited in their ability to self-assess and scheme out their behaviors, can manage to reach equilibrium. If social identity strength is subject to constant decay, and identity-affirming communication is stable and exerts a constant effect, then the strength of identity will naturally settle at whatever level that results from the combination of decay and affirmation. There are certainly some individual differences that will determine the rate of decay and the dose-response to communication, but speculating on the many possible causes of them is outside the scope of this project.

This has ramifications for how identity threat is managed. Those who have a strong identity and already have established a pattern of identity-affirming communication to maintain that identity have that communication to fall back on when identity threat occurs. Recall that identity threat first and foremost targets collective self-esteem. Negative collective self-esteem should only have spillover effects on strength of identification when it cannot be managed in some other way. The persistent use of identity-affirming communication helps to ensure a speedy return to one's norms in terms of collective self-esteem, thereby protecting the strength of identification. The weakly identified, who engage in less identity-affirming communication, are more likely to have to resort to reducing their group identification to manage threatened collective self-esteem. Some people will change their media use and social contacts to manage threat, but the stability of these constructs suggests these may not be the typical strategies.

What I am suggesting here is a tweak to the RSM. To the extent the theory has faced any criticism — at least in public — it has only to my knowledge focused on the problem of how stable media use and identity are (Scharkow, 2017, 2019). Although there are other ways to reason about or even question the basis of this concern, what I have suggested renders that potential problem essentially moot. By assuming identities have a natural tendency to decay in strength without affirmation, there is a theoretical basis for why communication can be unchanging in a dynamic system: it exerts a constant identity-reinforcing effect that counterbalances a constant identity decay. My suggestion about decay can also simplify the RSM, making it easier to consider RSM claims in the context of a single identity and without the need for enumerating threats. As currently constructed, RSM argues the need for identity maintenance is rooted in frequent identity threats — probably minor in severity — that exist in an open system. It also suggests one of the important countervailing forces that prevents positive feedback loops is the fact people tend to have multiple identities that may compete for time and have internal contradictions. I think it is correct that these play that role, but my slight reconfiguration makes these explanations no longer necessary conditions for the avoidance of positive feedback loops. One can assume that any time not spent maintaining a given identity comes at a cost to that identity. Specific threats to the identity can be enumerated, but it is not necessary in this simplified model.

This simplified theoretical model is visualized in Figure 1. Identity threat and time decay are formatted with dotted lines because they will not be directly observed — I will assume that shifts in collective self-esteem are due to identity threats and assume that identity strength is under pressure from time decay that must be counterbalanced by identity-affirming

communication. Identity-affirming communication has a direct effect on strength of identity, which is consistent with the usual media/communication effects approach. Such communication also serves to raise collective self-esteem, although this may only be noticeable in the presence of threat. Collective self-esteem influences strength of identity. The reinforcing aspect of the system comes from the reciprocal causation between identity-affirming communication and identity strength. There is ample opportunity for positive feedback loops within the three-variable relationship of identity-affirming communication, identity strength, and collective self-esteem. The exogenous forces of identity threat and time decay serve to counterbalance the reinforcement process.

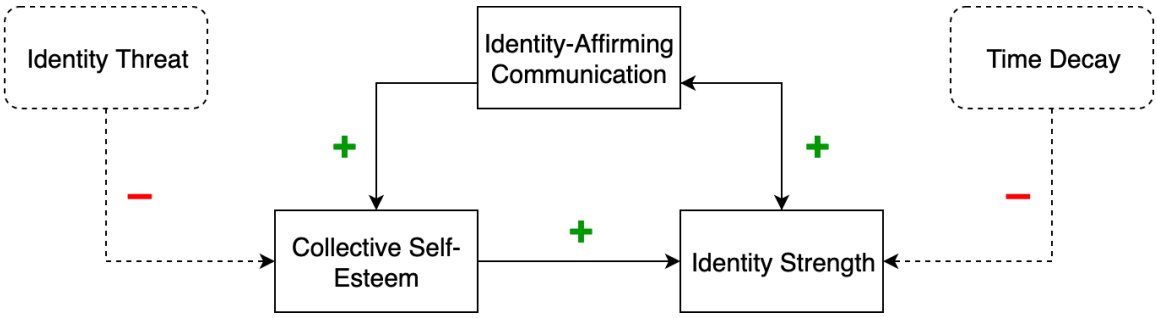


Figure 1: Simplified Model of Communication and Identity

Timing and Duration of Effects

The foregoing discussion raises questions about the temporal aspects of the causal processes described. Pinning down these details is challenging both theoretically and methodologically. As a general assumption, if a communication (mediated or otherwise) is to have an effect, it will happen almost instantaneously. It is not impossible to think of ways a change in attitude or belief may come after a delay, but in the rather non-specific context of this study, it is not be possible to enumerate or anticipate those cases. This assumption undergirds most experimental studies in which immediate post-treatment comparisons are the basis for conclusions about the existence of an effect. The more difficult question, and one that underpins examples in the previous section, is how long an effect of communication persists. Although I am not aware of research specifically assessing the timing of communication effects on identity, or the timing of the effects of identity on communication choices, there are two related literatures on duration of media effects on attitudes and the persistence of learned information in memory. The reason memory matters is that for the underlying attitude to be changed, an individual has to remember the new evaluation (although not necessarily the reason for the change or that the change occurred).

In a randomized field experiment in which televised political advertisements were the stimulus and (gubernatorial) candidate support the outcome, Gerber, Green, Gimpel, and Shaw (2011) found nontrivial impacts of the stimulus. But just a week post-treatment, effects were undetectable. Another field study of political advertising effects found substantial effects that tapered off to nearly zero within a week but could be sustained with frequent re-exposure (Hill,

Lo, Vavreck, & Zaller, 2013). After a laboratory experiment in which exposure to televised political incivility lowered political trust, Mutz and Reeves (2005) could no longer detect differences between experimental groups when re-contacted 3-4 weeks later. Hayes and Myers (2009) investigated the impact of the deaths of local servicemembers on support for the war in Iraq. Using local troop deaths in various timeframes to assess the timing of effects, they find that local troop deaths indeed reduce support for the war. The timeframe that captures this effect is troop deaths within the previous 2 weeks of the respondent interview, which may reflect some combination of the lag between troop deaths and local reporting and sample size issues in which there are sufficiently few troop deaths that many people live in an area in which a troop died in a given <10 day window. Most relevant is that the effects become undetectable on average by 3 weeks, suggesting effects that last about a week as well, on average (or most conservatively, up to 4 weeks).

To add to the complexity, there are clearly some important moderators of both the existence and persistence of communication effects. For instance, Bartels (2014) finds, like others, that persuasive impacts of a pro-Obama campaign advertisement undergo substantial decay over time. This decay was not experienced evenly, however, as effects endured among those predisposed to prefer Obama but not those who previously preferred Romney. Across domains, it should come as no surprise that the effects of messages may depend on pre-existing attitudes and identities (e.g., Comello & Farman, 2016; Garrett, Long, & Jeong, 2019). Similarly, the importance of the topic of a given communication influences the extent to which the message has an effect and how long that effect may last. Generally, greater persuasive effects are possible for issues that are of less importance to those exposed (Lecheler, de Vreese, & Slothuus, 2009).

When people care less about a subject, however, they put less effort into thinking about the message, consistent with dual-process models of persuasion (Chaiken, 1980; Petty & Cacioppo, 1986) and other findings (Chong & Druckman, 2010; McGraw, Lodge, & Stroh, 1990). Althaus, Bramlett, and Gimpel (2012) find impacts of local troop deaths on Iraq war support are much stronger and more persistent among those with relatively *low* levels of news exposure, suggesting those with high levels of exposure already had more fixed opinions, perhaps due to more knowledge. Knowledge has been shown to moderate duration and magnitude of effects in other contexts as well (e.g., Lecheler & de Vreese, 2011). Moreover, in naturalistic settings, the information environment tends to be crowded. Exposure to countervailing information, even in an unbalanced way, is to be expected. The typical observation in research is that exposure to two different views on the same issue in a short time period produces a canceling out effect, whereas if they are spaced out in time the most recent exposure prevails, but only for a short time (Chong & Druckman, 2010; Lecheler & de Vreese, 2013). This research does show more enduring effects when messages are processed carefully, likely due in part to further self-selected exposure (Chong & Druckman, 2010).

This review, of course, just scratches the surface. Dissertations can be and have been written on these topics. The present dissertation cannot incorporate the many complexities and must, as a simplifying assumption, treat the effects of communication as mostly homogenous across persons and relatively short-term. Some aspects can be probed, like partisan differences in effects and selection of communication. The proposed design will also offer the ability to explore the duration of communication effects in some more detail, but the prior research suggests effects will tend to be short-lived. This does not necessarily invalidate typical longitudinal

designs that space out interviews by weeks or months; it just means that low levels of communication are likely to have difficult-to-detect effects. That being said, for within-person statistical analyses, it will rarely be appropriate to use lagged measurements of communication because the effects are too proximate to the measurement of typical outcome variables. On another note, the temporal aspects of how changes in attitudes, identity, and so on affect communication choices are less clear. One potential contribution of the present study is to shed some light on whether the timing of the identity to communication selection causal pathway basically mirrors the reverse causal pathway, or not.

Chapter 5. Stability of Identity and Communication

The main threat to the validity and appropriateness of the aforementioned theoretical approach is that the key constructs are simply too stable to be able to demonstrate a causal relationship between them, if there is one at all. If identity, in particular, is extremely stable, then it will be difficult to link variation in any other construct with the meager over-time variation in identity. Although there is a great deal of evidence on the stability of the categorical aspect of identity, less exists in the case of strength and, to my knowledge, none has been done on collective self-esteem. High stability of categorical identity is not particularly problematic (and is analytically convenient) so long as there is sufficient variation in these other aspects of identity. Communication, on the other hand, could play this causal role regardless of its stability. My focus for this review is on the timescales these constructs are usually studied, with gaps between measures ranging from months to years. None of the person-level data collected for the research I will review was collected specifically to estimate the construct stability, so it is reasonable to think of the following as the stability of the constructs over the time periods in which they are commonly studied. These may not be ideal designs to study the theorized processes, however, because it almost certainly takes much less time than the lag period for the effects to occur.

Social Identity

I have enumerated several dimensions of social identity that are relevant to this context. One aspect of social identities subject to change is what is referred to as collective self-esteem (Luhtanen & Crocker, 1992), which refers to the valence one attaches to their social identities. Collective self-esteem is especially responsive to identity threat (e.g., Branscombe & Wann, 1994), at least in the experimental or cross-sectional contexts in which it is typically used. Although theory predicts some over-time variation in collective self-esteem, I am not aware of this measure having been used in longitudinal research. Besides using SIT to infer that it varies, I can only also look to abundant research on global self-esteem, the kind one normally thinks of when hearing the term self-esteem, to get an idea about variability in the collective dimension. Self-esteem research establishes both that the level of self-esteem is a trait-like quality. Mean levels tend not to change very much over long periods of time and measurement error-adjusted test-retest correlations (Heise, 1969) are approximately 0.7 for young and mid-life adults[6] with 3 years between measurements (Trzesniewski, Donnellan, & Robins, 2003). People who are relatively high or low in global self-esteem are fairly likely to stay that way in the long term.

On the other hand, a great deal of research has been conducted on the situational instability of self-esteem, showing that self-esteem indeed varies around its mean level and the amount of variability is a psychologically important variable (Kernis, 2005; Kernis, Cornell, Sun,

[6] For children under 12 and older adults (over 60), the stability measure drops to about 0.55. For adolescents, the stability measure is about 0.6.

Berry, & Harlow, 1993). Research on self-esteem stability typically occurs in the context of experience sampling studies with daily or similarly short intervals between measures (e.g., Updegraff, Emanuel, Suh, & Gallagher, 2010). To summarize, the evidence suggests that while the rank ordering of people's self-esteem is fairly stable, there is scientifically meaningful day-to-day variation. My expectation is that collective self-esteem will be no more stable than global self-esteem and quite possibly less, given the mechanisms described previously work to maintain global self-esteem when confronted with reductions in collective self-esteem. On the other hand, it is also plausible that collective self-esteem varies less than global self-esteem since there may be fewer and less variable influences of it. For those whose identity strength is low, it is likely they are less sensitive to threats as well.

There are at least two other variable aspects of social identity: strength (or intensity) of identity and the identity itself. Note that intensity of affiliation is not the same thing as collective self-esteem. A person can be deeply attached to a social group but still feel embarrassed, devalued, or a number of other negative (as well as positive) emotions about it. Reduced strength of identity may be a consequence of negative collective self-esteem, but is just one of several strategies available. What it means for the affiliation itself to change is more straightforward: The person no longer internalizes the categorization as part of the self-concept. It may be replaced with something else — a New York Mets fan could become a New York Yankees fan — but a new social identity is not a necessary component per se. I see the lack of identity as consistent with identity strength of zero. Some identities are in practice unchangeable, like race, but a person can vary in their strength of identity and in some cases may be able to internalize identification with other racial categories to varying extents, even those which do not correspond

to their family history. Unfortunately, research on identity strength and the adoption of identity groups is poorly developed relative to other parts of the social identity approach (Huddy, 2001).

Evidence suggests that the identities of greatest interest to social scientists are quite stable, at least in terms of the category remaining the same over time. By this I mean it is unusual for someone to change their ethnic identification, for instance, although it is not unheard of (Ethier & Deaux, 1994). The same research does show that shifts in the strength of these identifications are possible, especially during life transitions like adolescence and changes in geographic or social context (Sears, Fu, Henry, & Bui, 2003).

Categorical Partisan Identity

Politics is a useful context for considering the stability of social identities because partisan affiliation in places like the United States is both socially important and is far more malleable than racial or ethnic identity. A person can change their party identification at any time, but few do. Of course, shifts in the distribution of partisan affiliation can be quite consequential because partisans are so loyal when they cast votes (L. M. Bartels, 2000). To put some concrete numbers to this discussion, multi-wave panels conducted by the American National Election Studies (ANES) in the 1950s, 1970s, 1990s, and 2000s have been used to examine shifts in responses to a trichotomous item asking respondents whether they are Republican, Democrat, or independent. In these panels, which span 4 years, between 41% and 58% chose either Republican or Democrat in the first wave and did not change their response in subsequent waves (these numbers calculated by Clarke and McCutcheon (2009)). Between 8 and 21% chose independent in the initial wave and did not waver in their choice. Overall, then, consistent responses ranged from 57% to 69% of respondents across these studies.

This may seem like considerable variability, but it should be said those whose response changed were almost solely people switching to/from one party and independent, implying zero strength of affiliation with the former party in the conventional measures of party affiliation. People who chose both parties in the course of the panels ranged from a maximum of 7% of respondents to a minimum of just 4%. When considering those who "lean" to one party (they choose independent but in a follow-up say they are closer to one party than another) to be partisans, there is more partisan switching — about 12% switch from one party to another — but nearly 75% show no change, showing that many of those who shift from one party to independent still report leaning towards their former party. I must emphasize, however, that some amount of the change is surely an artifact of measurement moreso than true change in the underlying partisan attachment. The stability of partisanship is a long-running area of research within political science without clear consensus. Recent work suggests a large group of people, at least a majority of adults, have quite stable categorical partisan identifications in the United States, Canada, and United Kingdom but a substantial minority show a propensity to change as well (Clarke & McCutcheon, 2009). Empirical evidence shows categorical partisan identification to be fairly "sticky" — one whose identification changes at one point is likely to change back to their original identification — but changes among some populations are more often permanent and the overall evidence base is small (B. L. Bartels, Box-Steffensmeier, Smidt, & Smith, 2011).

Measurement of Partisan Identity Strength

In the United States setting, strength of identity is not typically well-measured in the many large-scale, public-use surveys available for researchers of politics. As discussed earlier, the measure in use at least since *The American Voter* (A. Campbell et al., 1960/1980) first asks

respondents if they consider themselves a Republican, Democrat, independent, or something else. If the respondent chooses Republican or Democrat, a follow-up asks whether they are a "strong" or "not strong" Republican/Democrat. Those who did not initially choose Republican or Democrat are asked a different follow-up about whether they "lean" towards one or the other party. A convention is to combine these into a single, quasi-continuous measure with the following values: very strong Democrat, not strong Democrat, lean Democrat, independent, lean Republican, not strong Republican, and strong Republican. The only problem with this face valid construction is that the measure has nonlinear relationships with many other constructs, like media use and political interest (Petrocik, 1974). The problem lies with people who initially call themselves independents, who are frequently quite loyal to one party and are relatively politically sophisticated compared to people who call themselves "not strong" partisans (Klar, 2014; Petrocik, 2009; also discussed in Long et al., 2019).

Bankert, Huddy, and Rosema (2017) introduced an alternative measure, based on social psychological measures of identity, for partisanship that generalizes to multi-party systems. It begins with a question and optional follow-up similar to the aforementioned, designed to get respondents to identify the party that they identify with most, even if they are only weakly identified. Once such a party has been identified, a multi-item scale is administered[7] tapping the extent to which the individual feels a member of the group. These validation studies have found

[7] Bankert et al. (2017) validated an 8-item measure following similar work in the U.S. using a 5-item scale by Huddy et al. (2015). Using psychometric analysis, Bankert et al. (2017) suggest a condensed 4-item scale for situations in which questionnaire space is at a premium.

that although this more elaborated measure does correlate fairly strongly with some cruder measures, it is more predictive of other constructs like participation.

Available data using these fairly new measures are limited. They were included, however, in the British Election Study (BES) panel that ran from 2014 to 2018, with 5 separate measurement occasions for this identity scale among a subset of respondents. Among those who maintained the same affiliation in the 4th wave (March 2015), 9th wave (July 2016), and 13th wave (June 2017), which comprised 82% of the respondents with complete data, the Heise (1969) stability measure is 0.86, suggesting high stability — this is among those who do not change their categorical affiliation at all, who are likely a more stable population than the whole. The interpretation of this number is that, after correcting for measurement error, the over-time correlation between measures of identity strength have a correlation of about .86. As one point of comparison, the stability coefficient for repeated administrations of IQ tests to adolescents reported by Heise (1969) ranged from .64 to .85. Unfortunately, how to make over-time comparisons with this measure among those who change affiliations is less straightforward since the scale items refer to a different political party.

Media Use

Unlike social identity, media consumption is a behavior. This is important because unlike a psychological attachment like identity, the gold standard for measurement of media use is not a self-report. There is little debate about whether self-reported measures of media exposure are accurate; they are not (Guess, 2015; LaCour & Vavreck, 2014; Prior, 2013a). Technological advances have made passive measurement sufficiently feasible to occasionally compare questionnaire measures of media consumption to observed consumption and such comparisons

invariably show self-reports to be quite inaccurate. The question that is debated is whether those self-reported measures are too wrong to be useful for some of the research questions they are used to pursue. Consistent over-reporting, for instance, does not render between- or within-subjects comparisons invalid even if the mean estimates of media use are wrong. More problematic would be if the accuracy or direction of inaccuracy is related to other key variables, but research has not yet demonstrated that this could be problematic for inference; for instance, responses to news use measurements do not appear to be related to the propensity to give socially desirable responses (Eveland, Hutchens, & Shen, 2009). Jürgens, Stark, and Magin (2019) point out that methods for passive tracking are hardly fool-proof as well, inducing bias in the samples available for this kind of tracking as well as documenting what appeared to be a technical error in a commercially available tool for media tracking that would only have been apparent on close analysis. There are simply no perfect options for media measurement. Although my focus in the research design will be specifically on political media use, especially partisan media, I will discuss what is known about media use more broadly in part because the issues are fairly general and because there is limited relevant work for partisan media usage patterns.

A complication is sorting out what exactly is being measured. Recent developments in measurement have addressed overly broad measures used in the past that made it difficult to determine what kinds of content people were exposed to. The "program list technique" (Dilliplane, Goldman, & Mutz, 2013) — which was designed for television but is just as easily ported to print, digital, and radio sources — presents respondents with a list of media sources. They are then asked if, over the course of a specified timeframe (usually a week or month), they have used the source regularly. The measure follows best practices identified by research on

questionnaire design: reports of behavior are more accurate when broken into their constituent parts (Menon, 1997), granular response choices can influence responses in difficult-to-predict ways (Schwarz, Hippler, Deutsch, & Strack, 1985), and regular behaviors are reported most accurately (Menon, 1993). Despite concerns that the number of programs does not correspond with amount of exposure (Prior, 2013a), validation research based on a sophisticated passive tracking system found a close relationship between the number of programs and total amount of media consumption (LaCour & Vavreck, 2014). The decomposition into separate programs also suits a wider variety of research problems, allowing for distinctions to be made about the kind, not just degree, of media consumption. A common use of this measure has been to differentiate consumption of partisan media sources from each and from non-partisan sources (Dilliplane, 2014; Long et al., 2019).

Given the existence of some measures suited to assess the stability of media use, how stable is it? Evidence suggests it is very stable, at least over the course of time that a typical panel study covers. In roughly a 6-month timeframe, Dilliplane et al. (2013) found a very high stability coefficient of over 0.9 for the program list technique using television programs. I used the same data (the 2008 National Annenberg Election Study) to calculate similarly high stability coefficients for the proportion of liberal (0.86), conservative (0.95), and non-partisan media (0.94) and the raw counts of each category are similarly high. Nevertheless, in a subsequent study Dilliplane (2014) found that changes, not absolute levels, in media use were most influential in changing vote choices and affect towards candidates in the 2008 election. This suggests the coarseness of the program list measure may not capture some change in media consumption until it becomes fairly significant or, perhaps, the Heise (1969) stability measure

overstates the proportion of variation that should be attributed to measurement error — a possibility I explore in a later chapter. Scharkow (2019) argues Heise's stability measure is inappropriately applied to the list technique as Heise assumes an interval measure whereas the list technique produces a sum of binary indicators. Nonetheless, a meta-analysis of other measures of television, internet, and print media (primarily news) used in 33 3-wave panel studies — mostly using hours/day or days/week frequency measurements — found an estimated stability coefficient of 0.9, quite consistent with Dilliplane et al.'s (2013) findings (Scharkow, 2019).

Although none of these studies cover the timespan that the 4-year studies of partisan identification do, the available evidence is at least suggestive that media use may be even more stable than party affiliation. This comes as a surprising finding if we assume it is true. There are some reasons, beyond prior expectations, to be skeptical of this interpretation. First, change in media use and change in social identity are being operationalized quite differently in the studies I have reviewed. Variation in party affiliation, as it has been studied by most political scientists, necessarily involves categorical change. It would be unusual to treat media use this way. One can conceptualize multiple categories of media use, but they are rarely mutually exclusive and usually vary in both degree and kind. Further, the kind of stability captured by the Heise (1969) measure concerns rank-order. In other words, the high stability of media usage measured by Dilliplane et al. (2013) and meta-analyzed by Scharkow (2019) is best interpreted as showing that it is unusual for a light media user to become a heavy media consumer and vice versa. Intra-individual change is basically ignored by this measure except when it changes the rank ordering

of respondents. These empirical results are also consistent with a model of media use varying in relatively short timeframes after which it reverts back to its mean.

This is important because intra-individual change in media use — a person differing from their norms — is quite often associated with theoretically expected changes in other variables. Fixed effects panel models, which completely disaggregate effects of changes in variables in a way that is robust to confounding (Allison, 2009), have shown changes in political media use can increase polarization (Dilliplane, 2014; Garrett et al., 2014; Wojcieszak, Azrout, & de Vreese, 2017), political participation (Kruikemeier & Shehata, 2017; Shehata, Ekström, & Olsson, 2016), political interest (Kruikemeier & Shehata, 2017; Moeller et al., 2018), and political expression (Gervais, 2014). When a large portion of the change in a longitudinal variable is due to measurement or some other kind of random error, then most of the variation available to model with fixed effects regression would be measurement error. The impressive pattern of results in these models that rely on intra-individual variability for statistical power in communication research suggest there is more than meets the eye when it comes to reports of high stability in media use. As I have suggested and will explore in more detail in a later chapter, the quantification strategies may overlook the more meaningful sort of variation. But if we accept the findings of high stability at face value, it suggests something that is perhaps not obvious. Fairly substantial effects as a result of changes in media use are observed despite its high stability, at least as stability is usually quantified. This is consistent with media being highly influential; that it is not more obvious owes only to the relative infrequency with which media habits change. Of course, this is just one interpretation. More granularity is also possible by

analyzing the content of media, which is likely to change more than media consumers change their selections (de Vreese et al., 2017), or by zooming in on day-to-day variations.

Social Networks and Discussion

One difficult aspect of describing social networks is the flexibility that exists in their definition and the uncertainty of how much the researcher's definition of the social network corresponds with a psychological reality (i.e., the person whose network it is thinks of himself/herself as part of the network). To somewhat narrow the problem, I will restrict this section to what is known about discussion and friendship networks. The former refers to networks in which a connection is defined as having had a discussion, either face-to-face or through some mediated channel. Usually the subject of these discussions are defined as whatever the subject considers "important matters" or politics, although I will emphasize these are not equivalent (Klofstad, McClurg, & Rolfe, 2009) but are perhaps more similar than one might expect (Eveland & Kleinman, 2013). One additional challenge with defining change in a discussion network is deciding how a tie is severed. Measurement can ease some of the operational burden, since a respondent can name the same discussants or not; conceptually, it is unclear how much time without a discussion must pass for the researcher (or the respondent, for that matter) to decide the connection no longer exists. Of most interest for the purposes of this dissertation is the partisan composition of the network, but it is theoretically useful to know the amount of fluidity in the network regardless of partisanship as well. The reason for this is that if people often change who they talk politics with, there are often opportunities to make these choices purposefully (e.g., to manage or avoid identity threat). Furthermore, although my hypotheses are not based on the partisan composition of the network *per se*, it is often the case

that this is the best surrogate available for the amount of identity-affirming discussion. The measurements used are, quite reasonably, limited in the number of discussants that can be elicited, making it likely that some of the heaviest communicators will report the maximum number of discussants at each opportunity and the only way to detect change is by looking at composition. Further, it appears the *amount* of discussion is another stable behavior per the typical measurements. To add some comparability with the previous discussion of media use, I calculated Heise (1969) stability estimates for frequency of discussion with a) family and friends and b) at work or online using data from the 2000 National Annenberg Election Study (NAES), which is the last time the NAES asked about discussion frequency to panel respondents on at least 3 occasions[8]. For political discussion with family and friends, the stability metric is 0.88 and for work and online discussion, the stability comes in at 0.91. This covers a timespan of about 4 months but does not deal with the consistency of the partisanship of these people.

There are some difficulties in assessing change via the standard "name generator" method, however. First, it is common for the sake of limiting response time that the amount of discussants that can be named is limited. Respondents with larger networks, then, will seem to have more variability if measured repeatedly; if they do not, then it suggests the assumption about the respondents' cognitive process to recall discussants produces non-random samples of the network, which may cause a different set of validity problems (Marin, 2004). Marin (2004) describes how these self-reports tend to produce measured networks that include a higher

[8] To be clear, I am referring to the 2000 NAES Multiple Reinterview Panel A. There were 672 valid responses to calculate the Heise (1969) quantities.

proportion of close connections than the actual network, which may not capture the type of ties that are likeliest to change over time. For instance, a great deal of research is concerned with discussions among people who hold different political views (e.g., Huckfeldt et al., 2004; Mutz, 2006). Recent work shows the standard name generator procedure likely understates the extent of discussion between those of different political affiliations as well as people of different races (Eveland et al., 2018), but shifts in the number or proportion of these ties are unlikely to be captured by standard measures. Note that, for my purposes, it may be a valuable tradeoff if these measures first elicit similar discussants.

One example of a study that is of a representative population of interest over a relatively short time period (6 months) is Bello and Rolfe's (2014) panel study of British voters in the run-up to national elections. Focusing only on those with whom respondents discussed politics, the authors state 53% to 59% of discussants re-appear in subsequent waves. One limitation of this study in terms of its comparability to the Mollenhorst et al. (2014) and Cornwell et al. (2014) studies is that respondents were not prompted to tell the researchers whether their second-wave discussants were indeed the same as those in the first wave, leaving it to the research team to make the connection based on initials and demographics of discussants and leaving out the opportunity for respondents to re-nominate forgotten discussants. Specifically, political discussions have a more instrumental (and specific) purpose, however, so it is plausible that there is more variation in general in terms of who is spoken to. Sokhey, Baker, and Djupe (2015) report on 3 separate 2-wave panels, each in a different country, in which respondents were administered name generators with 3- or 4-discussant maxima. The timing between waves ranged from 2 months to 4 years, but in the main the results are fairly consistent. They document

considerable variability in the reported networks; in the shortest interval between waves, just 53% of discussants were named in the second wave. With 4 years between measures, only about 28% persisted. As discussed, it is assumed that respondents perform some sort of memory sampling process, so the mere fact of not being named again is not proof-positive of removal from the network. But if random sampling was the only source of variability, there should be no difference between network variability between waves 2 months and 4 years apart. Illustrating how networks are both purposefully shaped while also being constrained by the social environment, Sokhey et al. (2015) show that like-minded discussants are more likely to be retained, they also find that the networks do not become more homogenous over time. What is not as clear is whether this could be the result of a process in which some people have networks that become more partisan and others whose networks become less partisan. In other words, it is not clear if networks do not (on average) become more partisan over time because very few people have changes in the partisan composition of their networks. The other possibility is that there is considerable over-time variation in network composition, but there are relatively equal amounts of change to and from partisan homogeneity. I have taken pains to show the amount of turnover in discussants because it suggests ample opportunity for this latter explanation to be true.

With those caveats, there is some relevant research on change within social networks. Wellman, Wong, Tindall, and Nazer (1997), examining whole intimate tie networks of a small number of Canadian adults over 10 years, find only about one quarter of ties persisted from start to finish. More recently, the National Social Life, Health, and Aging Project — a nationally representative panel study of Americans older than 57 — assessed change in the important

matters discussion network among that target population (Cornwell et al., 2014). In both waves, respondents were asked to name discussants. In the second, after the discussant nominations the respondents were shown a list of their prior-wave discussants and asked to confirm which were present in both lists. Follow-up questions probed dissolved ties. Overall, approximately 68% of discussants were retained over the 5 years of the study. Among the dissolved ties, around 17% were because the discussant had died and another 7.5% were because the discussant was too ill to maintain the relationship; these are not likely to be as common sources of network change among the general population. Another 15% of the dissolved ties were cases in which the respondent said they were still in touch, just to a lesser degree. Although we must be cautious in applying the findings among older Americans to everyone else, it is a very well-designed study that deserves consideration nonetheless. Using a similar procedure with a representative sample of the under-65 Dutch population, Mollenhorst, Volker, and Flap (2014) found a little over 50% of network members remained in the network after 7 years. Among the dissolved ties, around 20% were still in contact but at a frequency that was sufficiently reduced for the respondent to not want to re-nominate the first-wave discussant after prompting. Only about 7% of dissolved ties were due to death in these data. In both studies, the size of networks remained roughly the same although they do no formal measures of stability. An advantage of these studies is they do not impose a low upper limit on the number of network members.

Overall, there is clearly turnover in social networks, although on closer analysis some change is more of degree than in kind when considering the number of dissolved ties that are actually just less-frequent contacts. However, the studies I have highlighted tracing changes in social networks occur over 5 or more years, limiting comparability to the other focal concepts.

There are a number of studies documenting social network change in shorter time periods, but they tend to be on adolescents in school (e.g., Cantin & Boivin, 2004; Tanti, Stukas, Halloran, & Foddy, 2011) or people undergoing significant life changes that make maintenance of social networks difficult (e.g., Kalmijn, 2012; Leik & Chalkley, 1997). As far as the total amount of discussion is concerned, my analysis of NAES data suggest it is similarly stable to media use within the election season. Taken at face value, the research suggests that although the particular people do change, it appears people tend to maintain similar levels of both discussion and the partisan composition of the people with whom they discuss politics.

Chapter 6. Empirical and Theoretical Issues in the Study of Stability

The stability of a construct refers to how little it changes. Something that is completely stable does not change at all. Besides constructs that are immutable, we can treat stability as a property that varies continuously; some constructs are more (un)stable than others. It is not possible to engage in a nuanced discussion of stability without dealing with both operationalization and conceptualization. In the coming pages, my discussion of stability will be rather technical, but primarily in the service of having a better conceptual foundation. I begin with a simple description and then move on to a detailed explanation of the prevailing methods for measuring stability, which originate in research on measurement reliability. I show how these approaches have rarely-discussed assumptions that may lead to incorrect conclusions about the stability of constructs. Part of the reason for this detailed discussion is because a core threat to the validity of my proposed theory and research design is the possibility that the constructs under study are simply too stable to be studied. I argue up front for more care when making claims about the extremely high stability of a construct on the basis of the kind of evidence that is typically used to make such claims.

Quantifying the Stability of Constructs

Typically, to say a construct is stable is to put it in contrast to some reference standard. Such a standard is not often explicitly invoked, but the description "stable" implies the suitability of some types of analytic and theoretical approaches and the exclusion of others. For instance, in the analysis of repeated measures data, stable variables are considered a challenge because they do not exhibit enough within-unit variance to have statistical power (e.g., Clark & Linzer, 2015). Such claims are sensitive to the timeframe under consideration. As will be discussed in more detail later, media use is considered a highly stable behavior; yet in the course of a single day, the typical person will stop and start media use many times. It is only when aggregating at daily, weekly, or other levels that stability appears. Which treatment of time is appropriate depends on the research setting.

Assessing the stability of a construct requires measuring it repeatedly on the same subjects[9] over time. Armed with multiple measurements, the extent to which a subject's repeated measurements differ from one another is an indicator of stability. In applied social scientific research, however, one typically must distinguish measurement error from actual changes in the underlying construct. This assumes that the measurement $\mathrm{X_t}$ at time t is the sum of the true value T_t and random measurement error e_t:

[9] I will use the language of human subjects research, but this largely applies to other units of analysis.

$$X_t = T_t + e_t$$

This, as with most technical research on stability, comes from psychometric research on reliability (Lord & Novick, 1968). To assess reliability, the observed variance of X_t must be partitioned into random measurement error variance and true score variance:

$$Var(X_t) = Var(T_t) + Var(e_t)$$

The same information is needed to assess stability. It follows that

$$Var(T_t) \leq Var(X_t)$$

Although measurement and study design may aggregate out variance that would be observed at different time intervals (or timespans) as discussed earlier, a study design cannot support the estimation of more variance[10] in T_t than there is observed variance in X_t. As a logical extension of these relationships, any overestimate of error variance e_t will also overestimate stability. The reliability, denoted α^2, is defined in this framework, as

$$\alpha^2 = \rho_{XT}^2 = \frac{Var(T)}{Var(X)} = \frac{Var(T)}{Var(T) + Var(e)}$$

[10] This is not strictly true. In a "Heywood case" (Heywood, 1931), the reliability may be estimated to be greater than 1 (e.g., the measure explains more than 100% of variance of the true scores) and measurement error variance is therefore negative. Because these estimates cannot correspond with reality, I do not include them as estimates that could be treated as credible.

In plain terms, reliability is the proportion of observed variance that is due to variance in the underlying construct. The alternate notation of ρ_{XT}^2 signifies that reliability can be conceptualized as the (squared) correlation between the observed and true values. The more reliable the measurement X_t is, the more similar $Var(X_t)$ and $Var(T_t)$ will be. If reliability is underestimated, stability is overestimated. I point this out because much more research focuses on the proper estimation of reliability than there is research explicitly dealing with stability. One exception in which the reliability approach was used in a debate about construct stability was in political science, in which it has been argued that what was once described as incoherence (or non-existence) of political attitudes in the general public (Converse, 1962, 1964) was actually an artifact of highly unreliable measurements that are more consistent with quite stable attitudes once measurement error is corrected (Achen, 1975; Green & Palmquist, 1990).

To have enough information to statistically estimate $Var(T_t)$ and $Var(e_t)$, at least three t are required per subject (Heise, 1969). From here forward I will discuss these quantities as if they were obtained in a three-wave panel, but the approach generalizes to any greater number of measurement periods. Figure 2 shows a theoretical and empirical model connecting these concepts, which Alwin (2007, p. 105) calls the "quasi-simplex model" but appears without a name in other parts of the literature (Heise, 1969; Wheaton, Muthén, Alwin, & Summers, 1977; D. E. Wiley & Wiley, 1970).

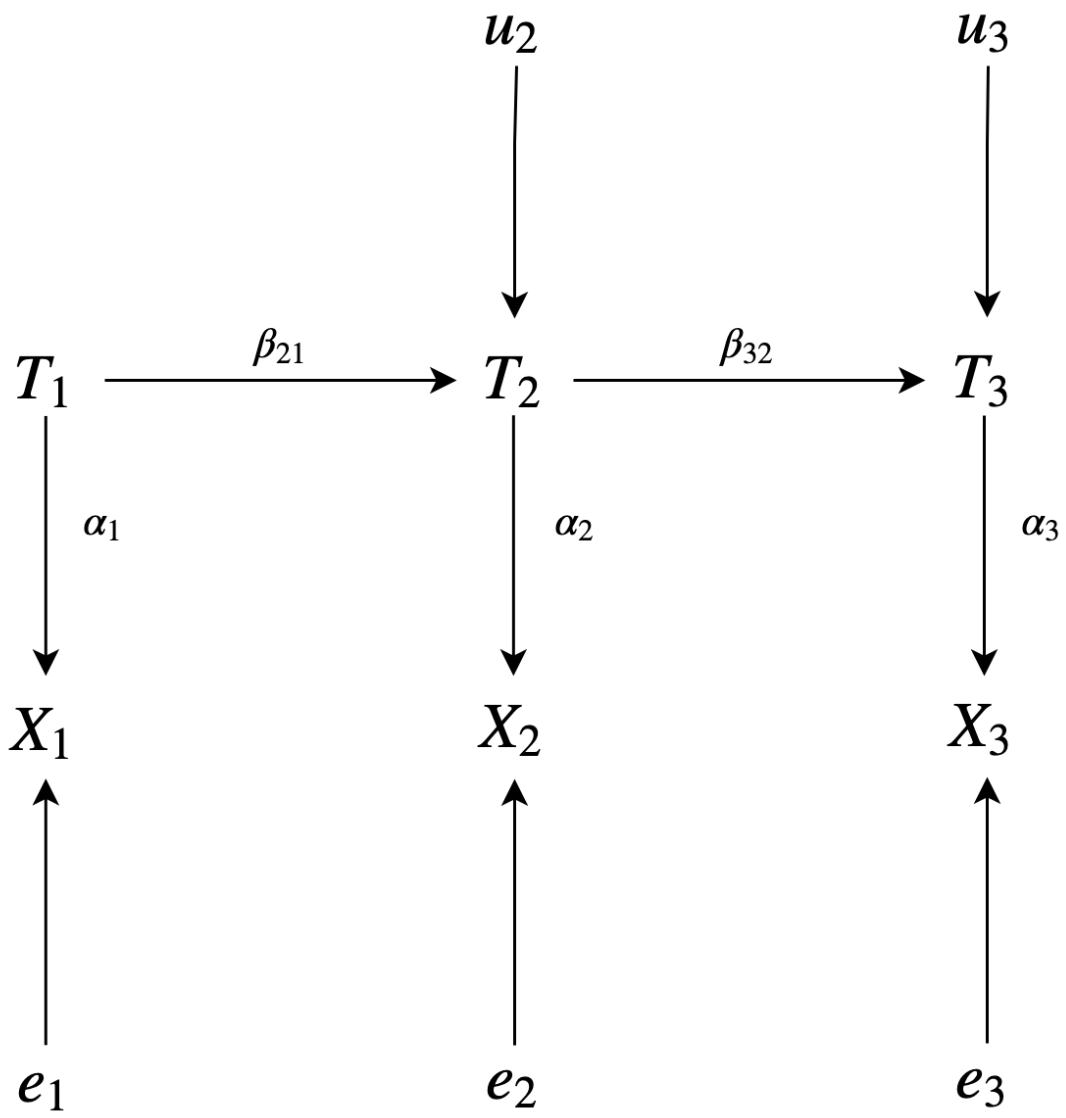


Figure 2: Quasi-simplex Model

In the quasi-simplex model, the true score at time t is understood to be determined by the true score at time $t-1$ as well as true score variance, $Var(T_t) = u_t$:

$$T_t = \beta_{t,t-1} T_{t-1} + u_t$$

$\beta_{t,t-1}$ represents one type of stability measurement. In the three-wave quasi-simplex model, β_{21} and β_{32} (see Figure 1) are two separate estimates of wave-to-wave stability. When standardized, $\beta_{t,t-1}$ is interpreted as the correlation between true scores from one wave to the next. The quasi-simplex model explains the observed correlation $r_{t,t-1}$ as:

$$r_{t,t-1} = \rho_{XT}^2 \beta_{t,t-1}$$

In other words, the observed correlation is the correlation between true scores as attenuated by reliability. This means $r_{t,t-1} \leq \beta_{t,t-1}$ except in Heywood cases. As my discussion proceeds, it may be useful to keep in mind that the practical boundaries for the true value of $\beta_{t,t-1}$ are $r_{t,t-1}$ at the lower bound and 1 at the upper bound. When I raise issues that suggest $\beta_{t,t-1}$ is often estimated in a way that is biased upward, the reader should bear in mind the true value cannot be lower than $r_{t,t-1}$.

The quasi-simplex model embeds several assumptions (see Alwin, 2007; Wheaton et al., 1977):

1. T_t is only influenced by its value at $t-1$ (*lag-1 assumption*)
2. There is no serial correlation in the measurement errors (i.e., $Cov(e_t, e_{t-p}) = 0$ for all p in 1 ... P where P is the number of panel waves prior to t)
3. True scores and measurement error are uncorrelated (i.e., $Cov(T_t, e_t) = 0$).

4. True score variance is not correlated with prior true scores (i.e., $Cov(u_t, T_{t-p}) = 0$ for all p in 1 ... P where P is the number of panel waves prior to t)

In the Heise (1969) quasi-simplex model, which is most often used by applied researchers to quantify stability, the reliability coefficients α_t are also constrained to be equal over time in order to identify the model[11]. A stability coefficient is estimated using equations in Heise (1969) that standardize the $\beta_{t,t-1}$ parameters to be interpretable as correlations and in the three-wave case, they are often multiplied together to yield a coefficient equivalent conceptually to β_{31}. This quantity, which can be computed using observed correlations, is fundamentally a measure of the rank order stability of subjects over time. That means if all subjects shift in the same direction and at the same magnitude over time, there is still absolute stability. An example of this is age — although age goes up as measurements are repeated, nobody becomes older or younger than anyone else relative to the first measurement. This alone would not affect stability estimates using the quasi-simplex model. Whether this captures the kind of stability one is interested in depends on the research question, but it is the primary if not sole measure of stability used in communication research (see Lee et al., 2008; Scharkow, 2019). For the time being, I take this approach on its own terms as a valid way to treat stability. Later, when I turn to quantifying stability at the individual level, I will consider the fact that there is more to stability than rank

[11] D.E. Wiley and J.A. Wiley (1970) instead constrain the error variances e_t to be equal, meaning the reliability can vary across time. These are two different assumptions that are not sufficiently important to explore in detail here.

order. There is also a case in which the interpretation of rank order-stability more closely matches the intuition of stability as lack of change. If the aggregate time series is stationary — that is, the aggregated mean and variance are constant over time, which is not extraordinary — then one can relatively safely interpret rank-order stability to mean lack of change.

A useful way to interpret the stability estimates $\beta_{t,t-1}$ (when standardized) is to consider their implications for analysis of the construct under study. Take for example a dataset in which the observed correlation of X from time 1 to 2 (r_{21}) and 2 to 3 (r_{32}) is .65 and the correlation from 1 to 3 (r_{31}) is .60. Using the Heise (1969) formulas, the reliability would be:

$$\rho_{XT}^2 = \frac{r_{21}r_{32}}{r_{31}} = \frac{.65 \times .65}{.6} \approx .70$$

Stability parameters would be:

$$\beta_{21} = \frac{r_{31}}{r_{32}} \approx 0.92, \qquad \beta_{32} = \frac{r_{31}}{r_{21}} \approx 0.92, \qquad \beta_{31} = \frac{r_{31}^2}{r_{21}r_{32}} \approx 0.85$$

Panel data regression models that focus on intraindividual change, like fixed effects regression and many offshoots (Allison, 2009), discard all between-subjects variance by design. If there is a high $\beta_{t,t-1}$ like 0.92, then only about 15% (i.e., $1 - 0.92^2$) of $Var(T_t)$ can be modeled. Whether this is a theoretically interesting amount of variance that is "up for grabs," so to speak, depends on the magnitude of $Var(T_t)$. Since T_t is not observed, we ultimately are modeling X_t. Given $r_{t,t-1}$ is 0.65, around 58% of the variation of X_t is subject to analysis but if the estimated reliability is correct, much of that variance is measurement error.

To show the consequences of underestimated reliability, let us suppose reliability is actually 0.9. We can then use the equality described earlier, $r_{t,t-1} = \rho_{XT}^2 \beta_{t,t-1}$, and solve for $\beta_{t,t-1}$:

$$\beta_{21} = \beta_{32} = \frac{r_{32}}{\rho_{32}^2} = \frac{.65}{.9} \approx .72$$

Although both 0.72 and 0.92 are reflective of fairly stable variables, the former figure is far more manageable from the perspective of statistical power (see Clark & Linzer, 2015).

The quasi-simplex model raises several theoretical questions. One is an assumption besides those I have already enumerated about this model: There are no other causes of T_t besides T_{t-1}, and if we relax the lag-1 assumption, still no other causes other than $T_{p<t}$. An alternate model, visualized in Figure 3, allows for another cause of T_t. In this model, adapted from Wiley and Wiley (1970), Y is a variable that also affects the value of T_t. For simplicity, Y is a time-invariant construct but it may theoretically also be something that changes across periods. Few are so naïve as to think that interesting social scientific constructs are actually caused only by their values in the past — some claim they rarely are caused by past values whatsoever (Achen, 2000) — so the question about how to specify the model again comes down to a mixture of empirical and theoretical considerations.

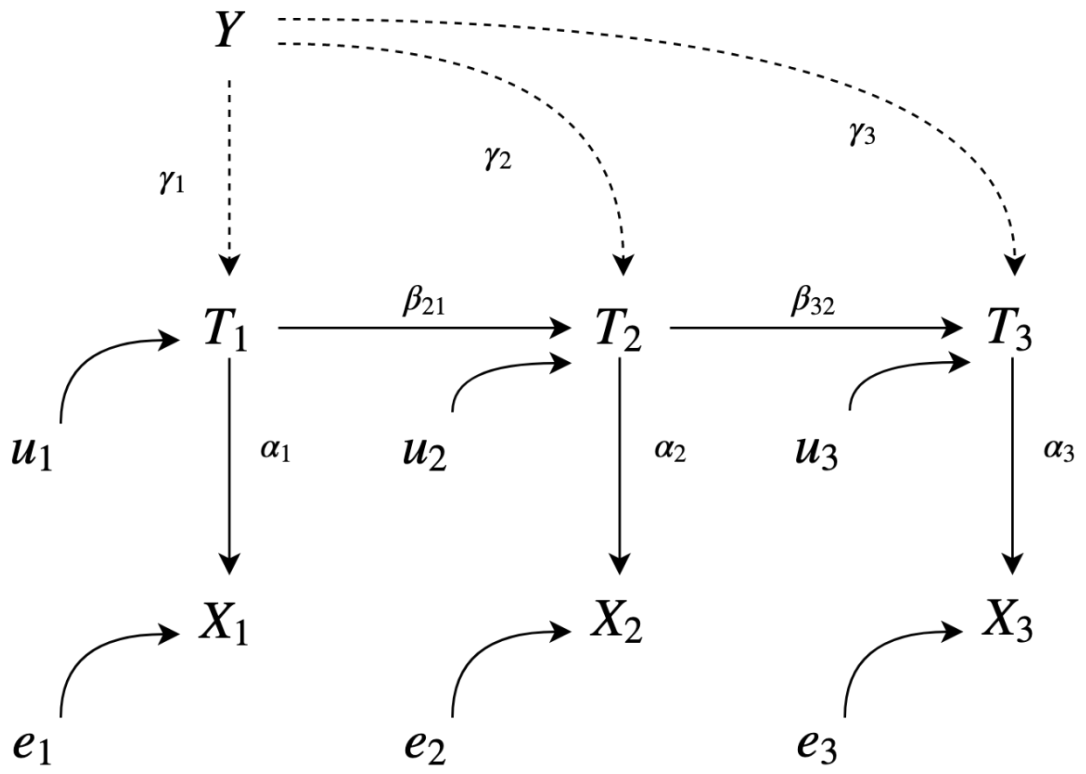


Figure 3: Stability Model with an Unmeasured Second Variable

In this case, the question is: what is stability? The quasi-simplex model and others like it are most useful when the interest is in identifying $Cov(T_t, T_{t-1})$ but there is no interest in u_t, the variance in T_t that is not caused by past values. This can be a valid simplifying assumption to make when the goal is simply characterizing stability while remaining agnostic about its causes. It is not an objective measure of $Cov(T_t, T_{t-1})$ because T_t is not directly observed — but it is a relatively simple approximation. Stability in this empirical-theoretical framework is what the over-time correlation of successive measures of T_t would be *if they were measured without error*. What is useful about a more fully-specified model, which will generally shrink the $\beta_{t,t-1}$ terms, is engaging in more complex counterfactual thinking. Using the multi-variable model in

Figure 2, the interpretation of the $B_{t,t-1}$ is what the over-time covariance of successive measures of T_t would be if they were measured without error *and the other causal variable Y was removed from the system*. This allows for T_t to appear stable but not be inherently stable. In other words, fully-specified models may lead to the conclusion that T_t is only stable due to the causal influence of other factors and would presumably not be stable if those causes were somehow removed or changed. This is my argument about the relationships between communication and identity. If the communication was removed or changed, the stability of identity would be lost (to some extent) and vice versa.

Although the decision to consider other causal variables in the measurement of stability is primarily a theoretical one, another potentially problematic assumption of the quasi-simplex model is more technical. The lag-1 assumption means that, in a three-wave model, T_3 is assumed to be

$$T_3 = \beta_{32}T_2 + u_3$$

Or expressed differently,

$$T_3 = \beta_{32}T_2 + \beta_{31}T_1 + u_3, \qquad \beta_{31} = 0$$

This formalizes the meaning of the lag-1 assumption: There is no added effect of T_1 on T_3 besides T_1's impact on T_2. When the true value of $\beta_{31} \neq 0$, it may not be obvious what impact this has on the meaning or accuracy of stability measurement. First, it may mean that the estimate of β_{32} is incorrect and that wave-to-wave stability is wrongly estimated. More importantly, as shown by Egan (2018), estimates of reliability are biased downward if the lag-1

assumption is violated. As shown before, downward bias in reliability means upward bias in stability since true variation is mistaken for random error. When the time period between measures is short, it is more likely this assumption is violated since whatever process that leads to autocorrelation between measurements may not take place on such a short timescale. For instance, if there are daily measures but the behavior occurs on a regular, 1-week basis, the 6th lag is the only one with a non-zero effect for the day the behavior occurs. Egan (2018) shows evidence the lag-1 assumption is violated for several social identity measures in a 3-wave panel in which measures are spaced 2 years apart, so it is not as if fairly long time lags are perfect protection against violations of this assumption. Properties of X_t likely to cause a downward bias in reliability estimates include relatively high wave-to-wave correlations (i.e., $\beta_{t,t-1}$) and low between-subject variance at wave 1 (Egan, 2018).

Why would applied researchers continue to use methods that can be biased under such common conditions? Besides the added analytic and conceptual complexity associated with relaxing the lag-1 assumption (and others already mentioned), three panel waves of a single indicator are insufficient to identify a structural equation model that includes the β_{31} parameter in addition to reliability estimates. A rough test of the lag-1 assumption is possible however, via estimating a regression equation of the form

$$X_3 = \hat{\beta}_{32}X_2 + \hat{\beta}_{31}X_1 + e_3$$

A value of $\hat{\beta}_{31}$ statistically distinct from zero suggests the lag-1 assumption has been violated. Unfortunately, it is not possible to both model the lag-2 effect and correct for measurement error. For the uses of the simplex model that I am most concerned with, however, the purpose is not to

correct for attenuation bias but to simply describe the stability of a variable. In this case, it is worthwhile to — at minimum — check this assumption as one way to judge the trustworthiness of reliability and stability estimates. Conceptually, it is more difficult to reduce stability to a single number once the lag-1 assumption is relaxed. As is always true in regression modeling, adding another predictor increases the explained variance in X_3. This means when the lag-1 assumption is violated, there is *more* variation in X_3 and by assumption T_3 explained by past values. This might seem to be an argument that conventional stability estimates when the lag-1 assumption is violated *under*estimate stability, then. More important, however, is that this increases the estimate of $Var(T)$ and decreases the estimate of $Var(e)$, the measurement error. The takeaway about violated lag-1 assumptions is that the true variation in the construct is underestimated. Using the regression procedure on the same NAES data from Dilliplane and colleagues (2013), I can demonstrate (details unreported for the brevity) that the lag-1 assumption is very clearly violated for all plausible operationalizations of media use and partisan media use — this casts doubt on the validity of the reliability and stability estimates in Scharkow (2019).

There are other issues that have ramifications for reliability estimation that I will not describe in such detail. Alwin (2007), for example, shows that for measures with fewer than about 16 response options, the use of Pearson correlations rather than polychoric correlations biased reliability estimates downward by about .1, which is far from trivial. The solution for calculating stability in the face of this problem is not so clear, however, because polychoric correlations cannot just be plugged into the Heise (1969) equations for calculating stability (Jagodzinski & Kühnel, 1987; Scharkow, 2019). The assumption mentioned only in passing

earlier about a lack of serial correlation in the errors may be untenable in many applied settings as well. The effect of serial correlation is more difficult to predict in part because it is an additional source of variance that is neither random error e_t nor true score variance u_t, but it stands to reason that unmodeled positive serial error correlations likely inflate estimates of stability since these errors are probably picked up by the $\beta_{t,t-1}$ terms in the simplex model. J. A. Wiley and M. G. Wiley (1974) propose an alternate model that models error correlations with different identifying restrictions, a model which Achen (1983) felt called into question his earlier claims of very high stability of attitudes (Achen, 1975). This model has not caught on to the same extent because it is suffers from a great deal of sampling error unless there are at least 4, and preferably 5 or more, measurement occasions (Palmquist & Green, 1992). Others still have discarded this general framework for longitudinal data in favor of approaches related to intraclass correlation focused on partitioning variation (Bland & Altman, 1996; Laenen, Alonso, Molenberghs, & Vangeneugden, 2009). All these models also assume linear relations, which is hardly an unusual assumption in applied research, but doing so is bound to conflate cyclic within-person processes like those expected in negative feedback system with random measurement error. I do not explore this problem in detail here because of its complexity and due to the high demands in terms of data to empirically differentiate cyclic longitudinal change from random error.

To review, there are a number of reasons to suspect estimates of stability using the quasi-simplex model may be biased upward. Although I have not proposed any simple alternative, I merely suggest that researchers have humility in the face of the considerable uncertainty involved in estimating measurement error-corrected stability.

Quantifying Stability at the Individual Level

The research question of primary interest for this dissertation concerns stability/variability of people, not constructs in aggregate. Aggregated construct stability is useful theoretically and pragmatically for making decisions about study designs, but we have to move to the individual level to learn about causes and individual differences. In this section, I will use stability and variability interchangeably, such that more variability means less stability and vice versa. It is often easier to talk about variability because this is sometimes what is really being measured; the lack of variability becomes what is understood to be stability. Most social scientific research in this area comes from psychology, where areas of focus include treating intraindividual variability in cognitive and other constructs as leading indicators of problems related to aging (e.g., Mroczek & Spiro, 2003) as well as intraindividual variability in affect and personality being related to measures of wellbeing (e.g., Greenier et al., 1999; Kernis et al., 1993). In any case, the end goal is to produce a single variability estimate for each person under study. It is possible, with sufficiently many observations of each individual, to construct a moving average variability estimate in order to model over-time changes in variability. For the present discussion, I will focus on producing just one estimate per person, but in all cases the researcher could restrict the "window" of observations to create a longitudinal series of variability estimates for each subject.

Ram and Gerstorf (2009) provide more nuance to the Nesselroade (1991; see also Nesselroade & Ram, 2004) development vs. variability distinction, separating types of variability they call *net intraindividual variability* and *time-structured intraindividual variability*. As the terms imply, net variability refers to the total amount of change without consideration of the time

ordering of the changes. The standard deviation is a way to quantify net variability; the quantity remains the same regardless of the order in which the observations occur. This means one could plausibly have the same standard deviation for two time series in which one is a straight line with non-zero slope and another that resembles an electrocardiogram (EKG). More substantively, net variability is independent of time in the sense that whether a deviation from the norm occurs at time t is not at all influenced by whether and how much there was a deviation from the norm at any previous times. Time-structured variability is generally considered to be the result of a dynamic process(es). An EKG has clear time-structured variability given that the level at any moment is contingent on the levels at several previous points in time. Whether a heart starts to beat has a lot to do with how much time it has been since the previous beat.

The likely most-used — and simplest — estimate of variability is the individual standard deviation (ISD), which deals primarily with net variability. I will consider measures of net variability first. The ISD is generally preferred over the individual variance because the ISD is interpretable on the scale of the variable under study. The downside, as mentioned earlier, is it does not differentiate between very different patterns of change. In particular, change-as-development increases ISD just as white noise does. One option to address this concern is to detrend the individual-level data before calculation, but in typical psychometric data this often induces artificial correlations between the detrended ISD and the individual mean scores (Baird, Le, & Lucas, 2006). Another suggestion has been to take either the raw or detrended ISD and divide it by the individual-level mean, yielding a coefficient of variation, but this has been criticized on the grounds of being an uninterpretable blend of the effects of variation, mean level, and the interaction between the two (Hultsch, Strauss, Hunter, & MacDonald, 2008; L. (Peggy)

Wang, Hamaker, & Bergeman, 2012). Another method that adds complication but avoids some of the most obvious issues of ISD and its derivatives is to use the residuals from a multilevel model with random intercept and growth curves (which detrend at the individual level) as the quantity of interest, which I will return to later.

Although there are many patterns that may describe the time-structured variation within an individual, I will keep the focus fairly straightforward. The time-structured equivalent of ISD — in terms of familiarity and ease of calculation — is the autocorrelation, AR(p), where p is the order of autocorrelation. For simplicity, I will discuss the AR(1) case in which the concern is autocorrelation between a value and its prior measure. AR(2) would mean the current value is affected by both its prior value and the value before that. Empirically, for each subject i there is a lagged dependent variable regression model:

$$Y_{it} = \alpha_i + \phi_i Y_{it-1} + \epsilon_{it}$$

If the variables are standardized, α_i drops out and ϕ_i is interpreted as an autocorrelation.

An apparent downside of estimating a correlation coefficient — or covariance, if it is not standardized — like this is that it is unusual to have enough observations per person to avoid considerable sampling variability. The reason it is apparent is because most researchers with conventional social science statistics training know that (for instance) 10 observations is a very small sample to use to estimate a correlation. That said, net variability measures like ISD have similarly undesirable small sample properties (Estabrook, Grimm, & Bowles, 2012). As in any

other case, the presence and extent of measurement error is an important influence on the accuracy of measures obtained from relatively few time periods.

Recommended by Wang, Hamaker, and Bergeman (2012) and developed over several subsequent publications (e.g., Jongerling, Laurenceau, & Hamaker, 2015) is what they and I will refer to as the multilevel AR(1) model. In the preceding, I described an estimation method in which one estimates a separate AR(1) model for each subject. This can be described as "no pooling" approach, in which the estimates for one subject has nothing to do with those for any other (in the terminology of Gelman & Hill, 2007). A pooled approach, in which the distinction between subjects is completely ignored, is clearly inappropriate. Multilevel models are a compromise between these two approaches that can aptly be described as "partial pooling" (Gelman & Hill, 2007). Using notation like Raudenbush and Bryk (2002), the multilevel AR(1) is

$$Y_{it} = \alpha_i + \phi_i Y_{it-1} + \epsilon_{it}$$

$$\alpha_i = \gamma_0 + u_{0i}, \qquad \phi_i = \gamma_1 + u_{1i}$$

Now we explicitly model the subject-specific intercepts and autocorrelation parameters as a function of γ_j, the population mean of the parameter j (the intercept or slope), and u_{ij}, the subject-specific deviation from the population mean of the parameter j. The u_{ij} are assumed normally distributed[12] with mean 0 and variance τ_j. Each u_{ij} have a covariance as well. By

[12] This assumption can be re-specified with any distribution provided there is software support.

modeling α_i and ϕ_i rather than mechanistically estimating them, the typical effect is that there is shrinkage applied to the estimates, towards the grand mean of each parameter. As long as the assumption that the subjects come from a common population is correct, the multilevel model makes full use of the information available by (in effect) looking skeptically at outliers. Although not the main reason one would use a multilevel model, this applies well in the case of unbalanced data (differing numbers of observations per person) such that estimates for those with less information are pulled more towards the population mean. In all cases, the more observations there are for a particular subject, the more the model will "trust" the non-pooled estimate.

The first equation is the level-1 (within-persons) model and is, in effect, a time series model. This means the caveats that apply to such models apply here, most importantly that the series be stationary (have constant mean and variance). At minimum, the lack of a constant mean can bias estimates. The simple fix for this is to detrend the series, which can usually be accomplished by including t as a predictor. Doing so makes the model equivalent to what psychologists call a growth curve model. Including a person-specific slope for time yields a latent growth curve model (Bliese & Ployhart, 2002; Hox & Stoel, 2005). Note that subtracting trends from the Y_{it} before modeling yields equivalent results to including trends in the model (L. (Peggy) Wang & Maxwell, 2015), so as a general recommendation I suggest modeling the trends since they may be theoretically interesting. Importantly, detrending avoids confounding change-as-development with the kind of variability that is the focus of this section.

In the multilevel AR(1), of interest is the ϕ_i, which is the person-specific autocorrelation. It is best to pause for a moment and reflect on the meaning of the autoregressive parameter. ϕ_i does *not* reflect rank-order stability; the rank ordering is effectively partialled out by the inclusion of the subject-specific intercept α_i and (if used) growth trends. Instead, the common interpretation of ϕ_i is as *inertia* (e.g., Suls, Green, & Hillis, 1998) or *regulatory weakness* (Hamaker, 2012). Greater absolute values of ϕ_i, then, are interpreted as being indicative of *in*stability. Why? First, consider the person-specific mean, α_i, as the equilibrium point. Note that this is part of the reason for detrending the series, because it enables an interpretation of α_i as the equilibrium net of any developmental processes (Nesselroade, 1991). α_i is also sometimes conceptualized as the trait component of Y_i, whereas the parts of the model that contribute to the estimation of Y_{it} are modeling the state, net of the trait. In a stationary series with $\phi_i = 0$, the value of Y_{it} conditional on α_i is ϵ_{it}, which is assumed to have mean 0. The residual ϵ_{it}, often referred to as a random shock, is assumed exogenous and by definition not predictable. The ϕ_i parameter reflects the extent to which ϵ_{it-1} is carried over to Y_{it}. That is, $Y_{it} = \alpha_i + \epsilon_{it}$ when $\phi_i = 0$. The expectation for Y_{it+1} in this case is just $E(Y_{it+1}|Y_{it}) = \alpha_i$, meaning we assume an immediate return to equilibrium. When ϕ_i is non-zero,

$$E(Y_{it+1}|Y_{it}) = \alpha_i + \phi_i\epsilon_{it-1} + \phi_i^2\epsilon_{it-2} + \cdots + \phi_i^{T-t}\epsilon_{iT-t}$$

where T is the total number of time periods. A positive ϕ_i means the random shock persists and a change to the underlying construct is expected to last into future periods with a magnitude of ϕ_i^p, where p is the number of time periods into the future. A negative ϕ_i means the system overcorrects the prior wave's deviation from equilibrium past the equilibrium point. Plotted over

time, a series with relatively higher ϕ_i will be characterized by broad peaks and valleys while negative ϕ_i produces rapid fluctuations around the equilibrium. Figure 4 shows three simulated time series in which the random shocks have mean 0 and standard deviation of 1, but with different autocorrelation values (ϕ). A useful metaphor is the path of the nose of an airplane. If random turbulence occurs, ideally the nose returns to its appropriate level as quickly as possible, as in the case of zero autocorrelation. High autocorrelation is like a pilot letting the plane veer off in the direction in which the plane was nudged. Negative autocorrelation is like the pilot constantly overcorrecting and essentially jerking the plane back and forth. Only the case of zero autocorrelation would be recognizable to the passenger as a stable flight path. Prior (2010) offers a similar interpretation of the ϕ_i parameter in his analysis of political interest, treating a near-zero ϕ_i as evidence for the very high stability of political interest through adulthood.

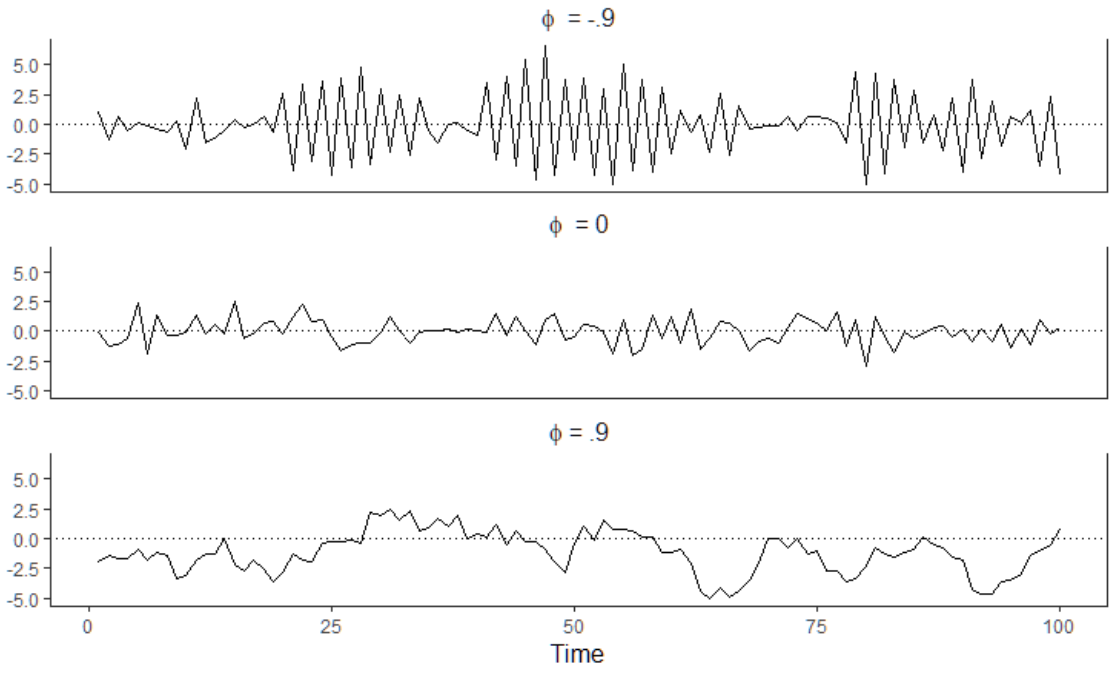


Figure 4: Time series with fixed innovation variance and different autocorrelations

To make statistical inferences about the correlates of ϕ_i, there are two main options: (1) fit the model, extract the ϕ_i, and run regression/correlation analyses on them or (2) incorporate the between-person predictors of ϕ_i into the multilevel model. Option 2 is what is known as a slopes-as-outcomes model (Raudenbush & Bryk, 2002). In a multilevel model, any predictor that does not vary within-person does not contribute whatsoever to the predicting deviations of Y_{it} from the person-specific intercept, α_i, because it provides no unique information for each t.

Instead, such a predictor only informs the estimation of α_i, the person-specific intercept[13]. In other words, the multilevel model with between-person predictor z_i

$$Y_{it} = \alpha_i + \phi_i Y_{it-1} + \gamma_{01} z_i + \epsilon_{it}$$

Can be re-expressed as

$$Y_{it} = \alpha_i + \phi_i Y_{it-1} + \epsilon_{it}$$

$$\alpha_i = \gamma_{00} + \gamma_{01} z_i + u_{0i}, \qquad \phi_i = \gamma_1 + u_{1i}$$

Which shows more clearly that the between-person predictor only factors into the estimation of α_i. In this specific case, the slope ϕ_i is of interest. Incorporating a between-person predictor works much the same way in that the goal is to get z_i into the ϕ_i equation. This is accomplished via an interaction term. The model

[13] A time-varying covariate will, generally, contribute both to the estimation of the person-specific mean α_i as well as Y_{it}'s deviations from it, which is not well-expressed by my notation. An exception is when all subjects have the same mean level of the time-varying predictor, in which case the covariates provide no unique information about α_i, only the Y_{it}'s deviations from it. This is, in fact, how so-called fixed effects models (also known as the within estimator) work in the multilevel framework. By subtracting the subject's own means from the time-varying predictor, all subjects have the same mean and therefore the variable's coefficient can only be interpreted as a within-subject effect (Allison, 2009; Bell & Jones, 2015; P. J. Curran & Bauer, 2011).

$$Y_{it} = \alpha_i + \gamma_{01} z_i + \phi_i Y_{it-1} + \gamma_{11} Y_{it-1} z_i + \epsilon_{it}$$

Expands to

$$Y_{it} = \alpha_i + \phi_i Y_{it-1} + \epsilon_{it}$$

$$\alpha_i = \gamma_{00} + \gamma_{01} z_i + u_{0i}, \qquad \phi_i = \gamma_{10} + \gamma_{11} z_i + u_{1i}$$

Which is known as the intercepts- and slopes-as-outcomes model (Raudenbush & Bryk, 2002). In sum, this approach allows for estimating the between-subject causes of over-time stability in the sense of time-structured variability.

An extension to this model allows for exploration of the between-person differences in the residuals, sometimes called *innovations* (Jongerling et al., 2015). An assumption of linear regression as well as multilevel linear models is homogeneity of residual variance. Multilevel models improve upon OLS regression by modeling the within-person correlation of residuals, but — at least in their basic implementation — still presume equal residual variance across persons. However, especially as the individual-level time series get longer, it becomes less believable that this assumption holds. Moreover, this is clearly an important aspect of variability regardless of ϕ_i. Even if there is no autocorrelation, high variance of the ϵ_{it} means the observed time series will be characterized by high variability. Figure 5 provides a simple visualization of the influence of the residual, or innovation, variance (σ^2) even when $\phi = 0$.

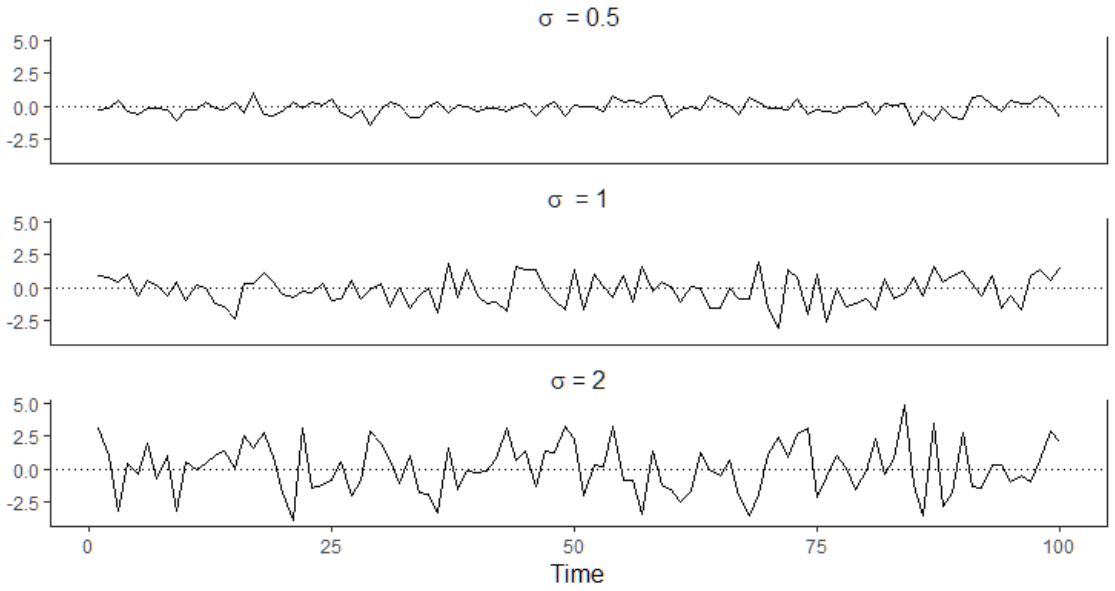


Figure 5: Time series with fixed autocorrelation and different innovation variances

Clearly, for a total description of stability, one must consider both ϕ_i, which corresponds most closely to time-structured variability, and σ_i^2, which corresponds to net variability.

Although some widely-used software for fitting multilevel models allows for person-specific residual variances, it is always treated as a nuisance parameter and including it as part of a model in which one predicts these variances with other variables is not feasible. One can sometimes extract these variances and analyze them outside the model, an option described earlier in the case of ϕ_i. It is strongly preferred, however, to estimate all the parameters within the model because they are interdependent (Jongerling et al., 2015). Bayesian estimation using Markov Chain Monte Carlo (MCMC) allows for models of arbitrary complexity and is particularly useful in the case of multilevel models in which there are many group-level parameters to estimate and relatively little information on which to rely for each group (Gelman

et al., 2014). I will spare the reader a full description of Bayesian statistical inference, but the main point is that it is indeed possible to use person-level variables to predict individual differences in residual variances. These models are essentially equivalent to what are referred to as multilevel models with heterogeneous variance (L. Hoffman, 2007) and "location scale" multilevel models (Hedeker, Mermelstein, & Demirtas, 2012) that are used for analysis of intensive longitudinal data in fields like sociology and public health.

To put these pieces together, we are left with two distinct types of variability that can be estimated simultaneously. The autoregressive parameter is a proxy for the extent to which the system tends to return to equilibrium following a change. As explained, the closer the autoregressive parameter is to 0, the more efficient the self-regulating process is. In my view, this is the best one-number statistical representation of the homeostatic process that characterizes the theory underpinning this dissertation. The residual variance reflects another important aspect of variability that is likely to have more to do with environmental factors and one-off occurrences. Although more difficult to statistically model, and for which relatively little is known about statistical power, residual variance is of clear importance in the more general context of variation and stability.

The multilevel AR(1) can be generalized to even more complex relationships, including and especially multiple (inter)dependent variables. Much as the multilevel AR(1) model generalizes single-level AR(1) time series analysis, these models can be the building block for multilevel vector autoregression (VAR; Bringmann et al., 2013; Schuurman, Ferrer, de Boer-Sonnenschein, & Hamaker, 2016) in which multiple time series are regressed on one another in a

way comparable to the cross-lagged panel model. These models can provide rich information about the over-time interdependencies of variables in a way that can be reasonably be interpreted as causal in addition to the types of insights I have just reviewed about modeling influences of stability.

Chapter 7. Theoretical Predictions

To clarify and reiterate the key predictions, I will express them as a series of statements. When I talk about relationships between variables, I am thinking at the within-person level. If X causes Y, this should — in my framework and in the design I will propose — be revealed as changes in X associating with changes in Y for each person. Or, in other words, if we had many observations of a single person, the relationships would still be in evidence even without other people for comparison. The usefulness of studying larger numbers of people comes from the fact that one expects people to be heterogeneous and we can average over the within-person results to account for both known and unknown contingencies in the causal processes.

To briefly review the preceding chapter, there are two types of change that may occur and I will have separate predictions for them. One is development, which refers to lasting changes to the mean-level of a construct. Although the duration of the proposed study will not be very long in the scheme of things, changes in development can be characterized as trends that persist across measurement occasions. The other type of change is variability, which has to do with the extent to which there are ephemeral changes in a construct of interest. As discussed previously, most communication effects measured in communication research seem to be more or less ephemeral, but with the ability to accumulate across time. Variability has to do with the magnitude and persistence of short-term changes in a construct, whatever the cause of the change. With these distinctions in mind, I proceed with predictions.

H1: Identity-affirming communication will predict decreased variability of identity strength.

Of course the RSM, from which I draw inspiration, suggests not only stability but reinforcement — that is, increases in identity strength — as a consequence of identity-affirming communication. As I have shown, it is possible to analytically separate variability from development. With this in mind, I predict:

H2: Identity-affirming communication will predict increases in identity strength.

In other words, identity-affirming communication promotes development. One thing to consider is that the relationship suggested by H2 could be contingent on identity strength itself. In other words, for someone whose partisan identity is weak, the predominant effect of identity-affirming communication may be to increase strength of identity rather than stability. This is consistent with my claim about decay. The weakly-identified are subject to little time decay, meaning the identity-enhancing effects of the communication should not be counteracted. Stronger identifiers, according to this logic, are likely to need identity-affirming communication just to retain a constant level of identity strength. I will propose this as a research question.

RQ1: Does the extent to which identity-affirming communication promotes variability and development of identity strength depend on the level of identity strength?

Next, I turn to collective self-esteem. Despite strong theoretical expectations, there is very little research exploring the causal relationship between communication, collective self-esteem, and identity strength in any combination. With that in mind, potential contributions of the present research include evaluating the following hypotheses:

H3: Identity-affirming communication will increase collective self-esteem.

H4: Collective self-esteem will be positively related to identity strength.

H4 is worded as it is because collective self-esteem is a bipolar construct and as such when it is low, I expect identity strength to be negatively affected and when it is high, I expect identity strength to be positively affected. Ceiling effects are plausible in both cases, however. H3 and H4 imply a mediation relationship as well, although my theoretical expectation is that identity-affirming communication enhances identity strength through more than just collective self-esteem. I conceive of the mediated path through collective self-esteem as fundamentally *protective*, in that this is how communication helps prevent reductions in identity strength. On the other hand, identity-affirming communication should give people reasons for continued identification with the group, information about the group and its qualities, and (at least in the interpersonal case) positive relational bonds with fellow group members that should affect identity strength without necessarily implicating or operating through collective self-esteem.

Thus far, my focus has been on aspects of identity as outcome variables because I expect them to be quite variable. Nevertheless, the population-level associations between identity and identity-affirming communication almost assuredly do not arise out of a process in which communication influences identity but the inverse is not true. As I have shown previously, that there are some indications — albeit not unequivocal — that communication behaviors are highly stable among adults when considered on the timespan of weeks, months, or years. If these findings are basically accurate, it may be the case that communication behaviors are determined largely by socialization processes and perhaps the periodic exogenous force (major historical events, personal crises, seemingly unrelated lifestyle changes) becomes the major source of

change thereafter. Even if this is the case, that does not imply the lack of a causal relationship even if it is difficult to identity. The communication can cause the identity to shift but the identity shift provides motivation to maintain the communication. In line with the assumed bidirectional relationship, I still hypothesize both directions of influence.

H5: Identity strength will predict decreased variability of identity-affirming communication.

H6: Increases in identity strength will predict increases in identity-affirming communication.

In line with the motivation rationale, H5 would be expected even if communication is very stable in general. A stable construct, as those who have studied the stability of communication conceptualize it, is one that seemingly randomly varies around its mean if it varies at all. Many everyday occurrences like holidays, social events, or weather may affect the feasibility or convenience of engaging in communication. As one becomes more strongly identified, it is likely the case that there will be more underlying motivation to persist in identity-affirming communication despite inconvenience. For instance, at Christmastime many cable news programs go on hiatus, use substitute on-air talent, or have different programs run in their place. The strong Republican may watch whatever is on Fox News on Christmas with his or her family while a less strongly-identified Republican might turn to seasonal, relatively apolitical programming before eventually returning to the his or her usual habits. The same logic can apply to interpersonal discussion, such as bringing up politics at family gatherings such as Thanksgiving, in which the most strongly identified may be more likely to carry on with political

discussion. The weakly identified, who may rarely engage in such conversations, may endure a temporary increase in this identity-relevant discussion (it may or may not be agreeable) before returning to the usual lower levels after the holiday. These examples can help to clarify the key distinction between identity and communication in that communication is a behavior that can be constrained by opportunity structures while identity is not intrinsically beholden to such external forces.

I have theorized collective self-esteem as affected by communication, subsequently affecting identity strength, but not affecting communication and not affected by identity strength. These predictions are theoretically defensible but not well-tested. With this in mind, I pose these research questions.

RQ2: Does collective self-esteem influence identity-affirming communication?

RQ3: Does identity strength influence collective self-esteem?

The most likely explanation if RQ2 is answered affirmatively is that people are more strategic than my model has given them credit for. They may selectively choose identity-affirming communication to manage their threatened self-esteem, suggesting more introspection than I expect. This relationship could be contingent, as in the fandom research in which people bask in reflected glory/cut off reflected failure (Cialdini et al., 1976) and do so differently depending on the strength of the underlying identity (Branscombe, Ellemers, Spears, & Doosje, 1999).

As for RQ3, it is less clear what to expect, if anything. Generally, the social identity approach as well as related theories like group vitality theory treat collective self-esteem as

something that varies independently of identity strength. It is partly an assessment of the opinion climate when there are no objective means (or too many to choose) by which to measure the group. It is not difficult, however, to imagine the strongly identified are more attuned to potential threats as well as reasons for group pride.

Last, I want to raise a question about identity-relevant, but not necessarily identity-affirming, communication. In politics, if we make the assumption that a program like NBC Nightly News is not partisan, then the expected impact is not clear in this framework. The strongly identified could plausibly receive it as threatening, given phenomena like the hostile media effect (Vallone, Ross, & Lepper, 1985) which has been shown to be contingent on the extent to which a person identifies with and perceives the source to identify with their political party (Reid, 2012). Of course, it should be noted that the hostile media effect does not manifest as consistently as may have originally been believed (Giner-Sorolla & Chaiken, 1994; Gunther, Miller, & Liebhart, 2009) and such perceptions need not necessarily be identity-threatening. Partisans also devote more cognitive energy towards refuting uncongenial arguments (Taber & Lodge, 2006), so mere exposure to potentially threatening information cannot be presumed to be accepted even if it becomes highly salient due to hostile media perceptions. The desire for exposure to nonpartisan programs, and presumably to engage in discussion with people of different or unknown partisanship, can be driven by political interest rather than just identity motives (e.g., Dubois & Blank, 2018). Maintaining interest also may as a byproduct increase identity strength through salience, especially since it can be difficult to follow politics (at least in the United States) without aligning oneself to some extent with a major party. Political events may dictate the effects of nonpartisan sources as well; good news cycles for the in-party may

make nonpartisan media have identity-affirming effects while bad news cycles for the in-party may cause nonpartisan media to have identity-threatening effects. Interpersonal discussion can be even more complicated at least from the researcher's vantage point since people get to know one another as individuals and may have expectations for how an out-partisan is "safe" for some topics and not others. As a catch-all, I will refer to political communication that is not identity-affirming as identity-*relevant*. Although I do not expect to resolve the causal ambiguity in my research design, it is quite likely I should be able to shed some light on the overall relationship. With this in mind, I pose the following broad research question.

RQ4: How do the effects of identity-relevant, but not identity-affirming, communication compare with those of identity-affirming communication?

Chapter 8. Intensive Longitudinal Study

To explore these hypotheses and research questions, I conducted an intensive longitudinal survey, in which I collected daily measurements of the key constructs over a period of weeks. As discussed in the preceding sections, relatively long series of repeated measurements compared to what is typically available in this research area are essential for gaining insight to the dynamics of identity and communication. The hypothesized effects are likely to occur on a relatively short timescale compared to the long time lags between measurements in traditional panel studies. When there is ambiguity about just how long it takes for an effect to occur or "wear off," it is better to measure more frequently than necessary than the inverse, since it is trivial to just aggregate the observations if it becomes apparent the measurement lags were shorter than the process dictates. The biggest tradeoff of this design is sample quality, as I used undergraduates recruited through the student research pool at Ohio State University's School of Communication. I argue that this is a reasonable tradeoff for access to such an intensive design, but it is a tradeoff nonetheless. One potential upside of a student sample is that it may exhibit higher volatility than usual when it comes to strength of political identification, thereby increasing statistical power to detect the hypothesized processes. Research on the RSM originated in part on research on highly-susceptible populations (Slater, Henry, Swaim, & Cardador, 2004) and this can be seen as an extension of that approach.

Overview

Participants were required to be U.S. citizens and at least 18 years old to ensure that they have sufficient stake in the U.S. political system and most institutional forms of participation

available to them. At first, interested respondents took an introductory survey that is approximately 10 minutes long. This initial survey allowed me to measure stable constructs as well as those that may be worth knowing but not to the extent that they need to be measured each day. In addition, the initial survey contained several more comprehensive measurements of the key concepts that are included in the daily surveys. After the first survey, participants filled out an approximately 2- to 3-minute survey with measures of identity strength, collective self-esteem, and identity-relevant communication each day for 20 days. Participants received encouragement to complete surveys each day regardless of whether they responded on the previous day or any number of previous days. With many measurements in a short time span, my prior expectation was that dropping participants with a single non-response would reduce the sample size dramatically — perhaps to zero. At any rate, missing data in such designs is common and retaining respondents with missing responses is considered a best practice (Ji, Chow, Schermerhorn, Jacobson, & Cummings, 2018). Other elements of the design, to be described in the coming pages, were designed with reducing respondent burdens as a key priority. In an intensive longitudinal design, if some aspect of the questionnaire is difficult or annoying to respond to, it does not just affect that questionnaire's responses but also the likelihood of receiving a response in subsequent days. This was such a priority, in fact, that the key measures of communication had to be devised *ad hoc* to suit this goal and therefore required development in a separate study.

Measurement Development Study

The measurement of communication, especially but not only of media exposure, is a central challenge to the field (e.g., de Vreese & Neijens, 2016; Eveland et al., 2009). In the context of the present project, the general challenges of measuring exposure via self-reports are compounded by an acute need to minimize the time and effort required to respond. Furthermore, the goal of capturing meaningful variation from one time point to the next requires measures that allow for a wide variety of responses. General best practice for asking respondents to report the frequency of regular behavior generally advises against allowing for precise estimates because such behaviors are not remembered in their particulars (episodically), but in terms of their approximate rate of occurrence (semantically), making it difficult to report the amount precisely (e.g., Menon, 1994). The solution in these cases is usually to give a coarse set of response options, like "not at all," "a little," "some," and "a lot." To elaborate, while a coarse set of response options with choices such as the aforementioned may be adequate to make distinctions *between* participants who are expected to differ from one another considerably in terms of their typical levels of communication, this project requires a measure that allows *within*-participant distinctions as well. For instance, suppose Participant A may range from 0 to 15 minutes of media exposure per day in a typical week while Participant B ranges from 3 to 6 hours per day in a typical week. With a coarse set of response options, one would expect Participant A to respond "not at all" or "a little" each day while Participant B would probably say "a lot" each day. This distinguishes the two from each other well but makes it difficult for Participant B to respond in a way that reflects the substantial variation in exposure.

The advantage of this project is that respondents will take surveys each day and can therefore be more reasonably be asked to estimate the actual duration of their behavior in the past day, because it can be remembered episodically. With this in mind, the following set of measures was created for potential use in the daily surveys:

1) **Yesterday,** how much did you talk (online or offline) about politics or news with people who…
    a) Are Republicans, support Republicans, or have a conservative point of view
    b) Are Democrats, support Democrats, or have a liberal point of view
    c) Do not support Republicans or Democrats and do not have a particularly conservative or liberal point of view

(Response choices: Open-ended numeric response in hours and minutes)

2) **Yesterday,** how much time did you spend reading/watching/listening to the news or politics content from the following types of sources, including articles and videos you've seen on social media:
    a) Sources that do not tend to favor one political party or ideology over another. Examples are *USA Today, Politico, Yahoo! News*.
    b) Sources that tend to favor the Republican party or conservative viewpoints. Examples include *FOX News, Breitbart News, The Daily Wire/Ben Shapiro.*
    c) Sources that tend to favor the Democratic party or liberal viewpoints. Examples include *MSNBC*, *Huffington Post, Mother Jones*.

(Response choices: Open-ended numeric response in hours and minutes)

These measures aim to be relatively easy to answer without much time commitment while still getting at the key constructs under study. Crucially, the project requires measurements of both media use and discussion, each of which must be subdivided into identity-affirming and not.

Although these proposed measures do not explicitly ask whether one's identity is in fact affirmed — seeing whether that occurs is a goal of the study and the reason for measuring components of identity separately — the measures tap into the extent to which the respondent perceives the communications to be from congenial sources. To evaluate whether these measurements can serve these purposes, I conducted a preliminary study using cognitive interviews.

Cognitive Interviewing Methodology

A cognitive interview, which is a 1-on-1, face-to-face interview between researcher and subject, involves asking the subject to respond to questionnaire measures and using a semi-structured interview to explore the subject's thought process in formulating a response. Such interviews are meant to assess the extent of comprehension, how memory is searched, and how these are synthesized into a questionnaire response (Willis, 2008). Of particular interest is identifying sources of response error and making changes to the measures to address them. I employed the *verbal probing* method (Willis, 2005), in which the interviewer asks specific questions about aspects of the response process rather than relying on the respondent to narrate all relevant parts of his or her own mental processing (Willson & Miller, 2014). Importantly, subjects are not asked to *evaluate* the measures, but are merely expected to describe what they understand them to mean and what considerations were brought to bear in coming up with a response.

Cognitive interviewing is a qualitative method and requires the interviewer to make on-the-fly judgments about appropriate questions to ask and the researcher — who, in the present study, is also the interviewer — to identify themes in the responses in a way similar to grounded theory methodology (Charmaz, 2014). And since the goal is to develop better questionnaire

measurements, decisions have to be made about making changes to the tested measures in the course of the study as well as whether enough subjects have been interviewed to feel confident in the final product.

Procedures

Overall, I conducted 16 cognitive interviews over the course of 3 days with subjects drawn from the same student research participation pool at Ohio State's School of Communication as used for the main survey study. Before the interview began, subjects were told the general purpose of the study — the need to see if the measurements are well-understood and capable of generating accurate responses. After this, subjects responded to the two measures under investigation using pen and paper. Once this was complete, the interview would begin. The following verbal probes were used in all interviews, typically in the order presented below:

- You said you [talked/used media] for [the time respondent provided] yesterday. Since we obviously don't usually time ourselves when we do these activities, how did you come up with this particular number?
    - [When all responses were 0 hours, 0 minutes]. Can you remember the last time you [talked/used media] and how long was that? [Then ask above question]
    - Since it isn't usually possible to report an exact time you spent doing these things, I know you are probably unsure about the time you gave. If you could give me a range of times, where the lowest number is the definite minimum time you spent [talking/using media] and the highest number is the definite maximum time you spent [talking/using media], what would that be?

- You said you spoke with people who [support Republicans/support Democrats/do not support either party]. Can you explain how you determined that this person or these people fall into that category?
    - [If hadn't talked to one of the categories] Can you remember the last time you talked to someone who [supports Republicans/supports Democrats/does not support either party]? [If yes, ask question above]
- You said you saw/read/listened to media that [tend to support Republicans/support Democrats/do not support either party]. Can you explain how you determined that the sources fall into that category?
    - [If hadn't used source from one of the categories] Can you remember the last time you saw/read/listened to a source that [supports Republicans/supports Democrats/does not support either party]? [If yes, ask question above]
- [If talked to anyone] When you talked to people about news or politics yesterday, was that in person, on the phone, online, or what?
- Do you use any social media sites or apps?
    - [For each] Do you ever see/talk about/hear about content about news and politics on there?
    - [For Facebook and Twitter, if used] If you see someone post about politics and it is not addressed directly to you, do you consider that talking for question 1 or more appropriate for question 2? And what if you respond, does that change the way you would categorize that time?

- You do not need to share any information you consider private, but can you tell me about what some of the subject matter was when you [talked/used media]?

In a number of cases, subjects would (in effect) answer some of the planned questions in the course of responding to another of the probes.

Results

Although the research goals of many of the questions are likely self-evident, there were several specific areas of concern I aimed to explore. The first, and perhaps most important, concern was the processes by which subjects came up with time estimates and whether these could reasonably be assumed accurate. Perhaps unsurprisingly, there was heterogeneity in the stated cognitive processes. Multiple subjects mentioned specific anchors they had available to aid estimation, such as one who looked at her digital calendar to refresh her memory about how much time there was in the passing period when she remembered reading news. Another had spoken to a relative on the phone and was able to look at his call log to determine the length of the phone call and use that as a basis for an estimate derived from his assessment of what portion of the call was devoted to news and politics. Yet another knew the entirety of his political media exposure was on YouTube from a specific channel whose videos are reliably of a specific length. Others had less or no external information available to self-verify their estimates, or at least were not interested in accessing such information. When asked to give a range of estimates to capture their overall uncertainty, respondents generally gave ranges that struck the interviewer as not particularly broad. Perhaps more noteworthy is the response was rarely the mid-point of the range of uncertainty, although this is not necessarily evidence of inaccuracy in the initial point

estimate. For example, one subject reported spending 5 minutes talking about politics and then provided a range estimate of 5 to 10 minutes. Another estimated 30 minutes, then provided a range of 20 to 30. The widest time range provided, in this case for media exposure, was from 20 minutes to 1 hour and 15 minutes, from a respondent whose initial point estimate was 30 minutes. Overall, the ways respondents described their strategies to estimate the time spent communicating were credible and likely to produce reasonably accurate estimates, albeit with imprecision.

Another area of emphasis in evaluating the measures was whether the respondents understood the distinction between the two broad categories of communication — talking about politics and seeing or hearing political content — in the potentially difficult cases of social media use. No respondent expressed any uncertainty about this distinction and agreed, for instance, that seeing content (both links to sources and one's own opinions) shared by people they know on a social media platform belongs in the media use category. All also agreed that direct interactions on these platforms would qualify as discussion. Examples of such interactions would be commenting on a Facebook post, replying to each other's tweets, and more direct methods such as private messaging and chats. These understandings are in line with researcher expectations. An additional categorical distinction explored in the interviews regards the partisanship of sources. When it comes to discussion, respondents generally expressed confidence in the partisan leanings of the people to whom they talk. When asked how they come to these conclusions, most commonly respondents said it was apparent from the opinions expressed on policy issues. In a few cases, the signals were far clearer, like a respondent who said he spent time each day in a group chat among members of his local College Republicans group. Respondents were more

uncertain when it came to media sources, particularly in terms of identifying sources that are non-partisan. Some expressed a belief that all sources are biased one way or another. The interviewer did not notice any clear category errors (e.g., FOX News is non-partisan, BBC supports Republicans) when probing about specific sources, although of course reasonable people may disagree in some cases. In this sample, social media platforms comprise the major source for political information and news, which typically includes both traditional media and other people expressing their opinions. Respondents indicate that they usually know the partisan preferences of people they follow on social media sites, either from their offline relationships or by noticing patterns in the things posted online. Although these measures will not allow for a clear idea of the particular sources respondents are exposed to, and there will be cases in which the same source may be categorized differently across respondents, this scheme appears suitable for the purposes of the study.

Finally, some probing was devoted to exploring which kinds of topics were considered "news and politics" by respondents. Prior work suggests heterogeneity across people in their definitions of what is political (Fitzgerald, 2013; Morey & Eveland, 2016), which can introduce non-random measurement error. Although some heterogeneity was apparent, no particular patterns were apparent. The inclusion of the term "news" in the question wording appeared to expand respondents' definitions of what topics were relevant. Most topics respondents mentioned would fall under any reasonable definition of politics; edge cases are topics more relevant at the sub-national level, like crime news, which are appropriately political in this researcher's view.

Revisions to Items

As expected, the interviews revealed some shortcomings in the measures as originally constructed. One relatively minor change is in the ordering of the partisan subcategories, which were not consistent across the two measures. Furthermore, given the extent to which respondents seemed to feel that "news" was a broader category than "politics," the first item's prompt was altered to read "news or politics" rather than "politics or news." Most significant is a change to the response format. The original measures ask respondents to report the time they spent on each form of communication separately for non-partisan, Republican-favoring, and Democrat-favoring sources. The goal behind this response format was to simultaneously allow for granularity in response without prompting respondents to think explicitly about the ratio of in-party vs. out-party communication, which could be susceptible to desirability biases. However, respondents typically explained their estimation processes as starting with an estimate of the total time spent communicating and then an estimate of the proportion of time with each partisan category. To adapt the items to the respondents' cognitive process, they are asked first to give an estimate of the total time communicating and then the proportions of time with each partisan category. The final items are the following:

- **Yesterday,** how much did you talk (online or offline) about news or politics?
  [ ] hours [ ] minutes
    - And about what percentage of that time was with people who…
      - Are Republicans, support Republicans, or have a conservative point of view
        _____ %
      - Are Democrats, support Democrats, or have a liberal point of view
        _____ %

        - Do not support Republicans or Democrats and do not have a particularly conservative or liberal point of view

          _____ %

- **Yesterday,** how much time did you spend reading, watching, listening to, or hearing about the news or political content, including posts you saw on social media?

  [ ] hours [ ] minutes

  - And about what percentage of that time was the content from sources that…
    - Sources that tend to favor the Republican party or conservative viewpoints. Examples of sources like this include *FOX News, Breitbart News, The Daily Wire/Ben Shapiro.*

      _____ %
    - Sources that tend to favor the Democratic party or liberal viewpoints. Examples of sources like this include *MSNBC*, *Huffington Post, Mother Jones*.

      _____ %
    - Sources that do not tend to favor one political party or ideology over another. Examples of sources like this include *USA Today, Politico, Yahoo! News*.

      _____ %

The final 10 interviews evaluated these revised questions and did not reveal any new problems specific to them. As a note of clarification, the technical implementation of these items in the survey forces the percentage responses to sum to 100% whenever participants report greater than zero communication.

Main Study

Data collection began on January 30th, 2020 and analyses reflect those who completed their participation by March 14th, 2020. Readers should note that Ohio State's spring break began on March 9th and on March 10th, the university announced the suspension of all in-person classes until the end of that month due to the novel coronavirus (COVID-19) pandemic. The surveys were programmed and distributed via "formr," a survey design and distribution framework created to accommodate complex designs like this one (Arslan, Walther, & Tata, 2019). formr hosted the questionnaires on the web — with access restricted by personalized links that are sent via email and text message — and automated the process of re-sending questionnaires at pre-specified intervals. After taking the introductory survey, respondents were sent an email each day at 8:00 AM containing a link to that day's daily survey along with an update on their current progress in the study (how many days they have participated, how many remain, and the number of credits earned). Responses had to be entered before midnight or else are considered missing for the day. Those who provided a cell phone number in the introductory survey were also sent a text message each morning with a link to take the survey[14]. The goal was to have the survey filled out in the morning, ideally before any political communication. If the communication measures referred to the *present day*, and were administered later in the day, it would be difficult to pick a time that respondents are both likely to be done talking about politics and

[14] This was managed with the aid of a service called Twilio, which allows sending these messages from a local number used exclusively for this study.

watching/reading/listening to content about politics *and* apt to respond before the day is out. The chosen strategy risks having effects dissipate overnight, but should better minimize non-response and under-reporting. The median time to complete the daily surveys was 81 seconds, while the 10th percentile completion time was 43 seconds and the 90th percentile was 215 seconds. The mean of over 9 minutes is greatly influenced by a handful of cases in which respondents apparently completed the survey long after visiting the webpage. For the introductory survey, the median response took 8 minutes with an approximate range of 5 to 15 minutes.

Statistical power for the analyses to be conducted involves two key dimensions of the design: The number of participants and the number of measurements per participant. Prior research shows that for typical within-subjects statistical models, there are major gains in statistical power available by increasing from the standard panel lengths (usually 2-5) to 10, 20, or more (Clark & Linzer, 2015)[15]. This is particularly true if the variables of interest do not change a great deal, which very well may be the case for the present study. For the simplified form of the statistical models of interest, Jongerling and colleagues (2015) find reasonable performance with N = 100 subjects and T = 10 measures; I chose T = 21, although some missingness was expected — and observed — and lagging variables will cost at least 1 observation per subject. Because of a technical error related to the time zone difference in

[15] For instance, for variables that are relatively stable, Clark and Linzer (2015) find that going from 5 to 20 measures reduces the root mean squared error of regression coefficients by half with 100 subjects.

Europe where the formr server is hosted, participants were sent 22 surveys rather than the 21 originally planned (and incentives were adjusted correspondingly). I did not impose a ceiling on the number of subjects who could enroll, but I targeted a minimum of 100 and ultimately ended up with 216 who were analyzed after exclusions. That being said, I used a slightly more complex version of the model than the one assessed in that study which likely comes at some cost to statistical power. Schultzberg and Muthén (2018) used simulation-based testing on a closely-related model to the one I used and found that for most parameters of interest, an N of 200 had very high power with T > 10. They conclude that overall, it is "clear" that "large N seems able to compensate for small T" (p. 511). That statement comes out of an intensive longitudinal design context where even T = 20 is considered small and N > 100 is considered large. The main parameter that may have poor statistical power in this design due to too few observations per participant is the residual variance, an aspect of quantifying variability. Although the number of observations per subject is at the low end of acceptable statistical power by some measures, I regard any longer series of measurements as outside of what can be facilitated within C-REP or any other cost-efficient option.

Missing Data

As previously mentioned, an expected feature of these data is a high level of missingness. Without expensive incentives, most participants will not respond each day. Although the deletion of missing data is rarely advised in quantitative research, the problems are arguably more acute in longitudinal designs when the use of lagged variables make such data losses compound over time. To address the missing data, I use the multiple imputation procedures implemented in the Amelia software package (Honaker, King, & Blackwell, 2011) which has several features

designed specifically for longitudinal data that allow the imputation procedures to consider time trends, lagged values, and so on (Honaker & King, 2010). A full explanation of multiple imputation is beyond the scope of this project, but the basic procedure is that a statistical model makes probabilistic predictions about what values would have been observed if the respondent had taken the survey and imputes them into the data. To account for the uncertainty in these probabilistic predictions, multiple imputed datasets are created with different plausible values for the missing observations. Analyses are then run on each imputed dataset and the results of the analyses are combined after the fact to capture the variation in potential results. Although for simple analyses only a small number (e.g., 5) of imputed datasets are necessary, larger numbers are suggested for Bayesian estimation (e.g., 20 to 100; Zhou & Reiter, 2010). The analyses reported here are based on 25 imputed datasets. Appendix A2 includes visual comparisons of model results using multiply imputed and complete data.

The main causes of missing data are skipped surveys and attrition. Recent research suggests that skipped surveys should be imputed using multiple imputation procedures (Ji et al., 2018) and dropping participants who fail to meet a certain level of participation can in some cases be worse than doing no accounting for missing data at all and analyzing only the observed cases (Jacobson, 2020). There are few guidelines or clear justifications available in the literature for choosing a threshold of completion under which all participant data should be dropped (Trull & Ebner-Priemer, 2020), but the findings of Jacobson (2020) indicate that more inclusive standards are best. Ji and colleagues (2018) used simulations to find satisfactory performance without any threshold and 30% of all observations across participants missing. Jacobson, Chow, and Newman (2019), using a related modeling approach to mine, find strong results even when

70% of all observations are missing. The goal of the missing data procedures, then, is to not omit participants who participate less since their non-participation may be related to important variables. If those participants are not informative, then they should just add noise to the statistical estimates. With this in mind, all participants who completed at least 3 surveys were included in analyses and all their missing responses were imputed. Appendix A2 includes a visual comparison of regression results when different inclusion cutoffs are chosen. These exclusions shrunk the final sample from 262 to 216. It should be noted that recruitment materials made clear to participants that participation beyond the initial survey was not required. After the exclusions, just under 35% of all possible observations remained missing and were therefore imputed.

To give an overall impression on the prevalence and patterns of nonresponse, Figure 6 is a "missingness map," a visual overview of the missingness in the data (Alemzadeh et al., 2020). In it, each row represents a single respondent and each column represents a survey. The color of the cell indicates whether the respondent completed the survey. It is sorted from highest to lowest response rate to show the point at which participants were excluded. Given the lack of prior research addressing some of the particulars in terms of the decisionmaking regarding exclusion of participants and which missing observations to impute, I have done a small-scale simulation study detailed in Appendix A3 exploring the consequences of several researcher decisions. Put briefly, results indicate no obvious best choice, but excluding participants who do not participate at a high level has many of the same costs as complete case analysis and few of the benefits of multiple imputation (e.g., statistical efficiency). When there is non-ignorable missingness related to the focal variables, the procedures used here appear the most robust and efficient.

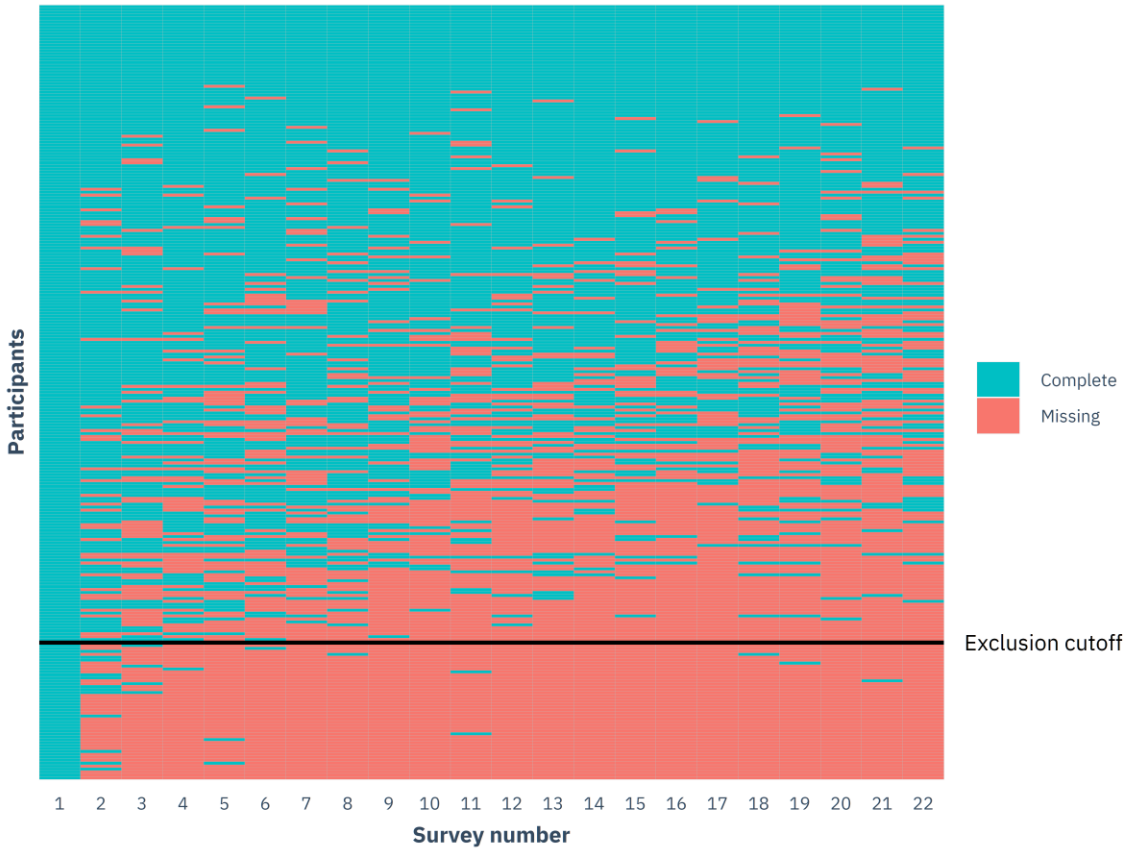


Figure 6: Missingness Map

In Figure 7, the missingness map is plotted over time to show the distribution of survey responses over the course of data collection and to show that there does not appear to be any pattern of missingness that corresponds to which of the approximately 6 weeks respondents were participating in.

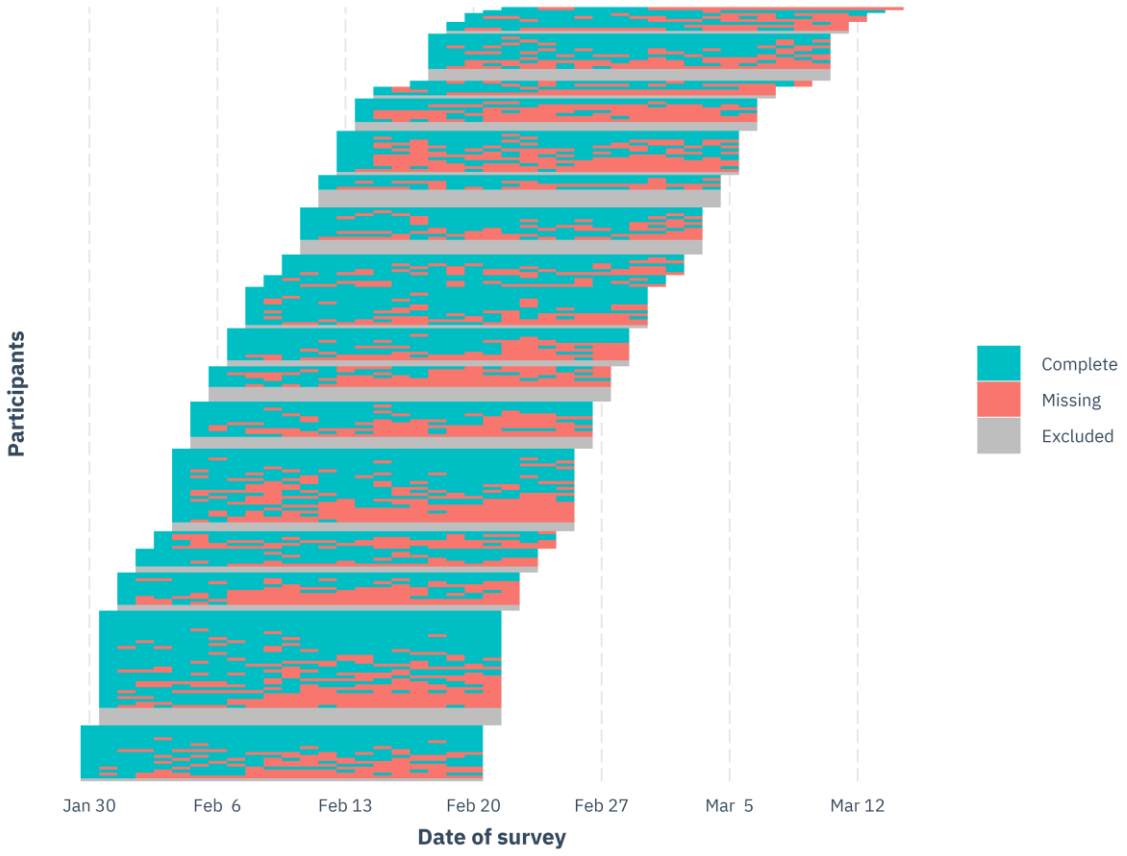


Figure 7: Survey Response and Nonresponse Over Time

Although the study design is inherently restricted in its ability to generalize to more diverse populations, some simple analyses were conducted to explore the potential consequences of the missing data. First, I created a simple logistic regression model using predictors from the initial survey to predict whether respondents would clear the bar of 3 responses to be included in the analyses. Using age, race and ethnicity, partisanship, whether the respondent voted before, self-reported attention to politics, issue ideology, the number of days in an average week the respondent reported talking with others, the number of news sources the respondent reported

using, and identity strength, the overall model fit failed to perform better than an intercept-only model, χ^2 (11) = 19.25, p = 0.06. Although it is probable those who did and did not respond often enough for inclusion differed in some ways relevant to the study, these differences do not appear to be large. Next, analyzing only those who were included in the analyses, I fit a Poisson count regression model predicting the number of incomplete responses using the same set of predictors. The omnibus test for this model was statistically significant compared to a null model, but no individual predictor variable was associated with significantly more or fewer responses. This suggests that indeed there are differences between the highest and lowest responders, but it is not possible to characterize them very specifically. This does not capture all potentially important causes of missingness, but suggests that the causes of nonresponse were not primarily between-subjects differences in these key variables.

Measures

All question wordings are included in Appendix A1 and were presented in that same order in the daily questionnaires (discussion, media, identification, collective self-esteem). Respondent characteristics like age, gender, and so on were collected in the introductory survey. In that first survey, respondents responded to a conventional categorical party identification item. First, they were asked, "generally speaking, do you think of yourself as a…" with response options "Republican," "Democrat," "Independent," and "Something else." Those who choose "something else" (or skipped the first question) were given a forced choice follow-up asking, "Do you generally think of yourself as a little closer to the Republicans or Democrats?" Those who did not choose one or the other were notified that their continued participation required a

choice. The final sample was composed of 69% Democrats and 31% Republicans. 38% of respondents initially chose “Independent” or “Something else.” All descriptive statistics refer to the 216 respondents who supplied sufficiently complete data to be retained for analysis and only the complete, non-imputed responses. For measures that are repeated, the descriptive statistics refer to the average of respondent averages; that is, a respondent with 22 responses is not weighted any more than a respondent with only 3 since the statistics are calculated based on each of their own mean responses.

After this, and on each of the daily surveys, partisan strength was measured by items adapted from the partisan identity scale developed by Huddy and colleagues (2015; Bankert et al., 2017). Unlike the validation studies for this measure, I ensure that respondents receive measures for the same identity throughout the entirety of the study, a decision more justifiable given the short duration. To reduce memory effects as well as the length of the questionnaire, only a random subset of the items for the identity strength questionnaire was administered each day. This is a form of planned missingness, a procedure designed to reduce respondent burden and in my case increase validity in narrowly-spaced repeated measurements (Silvia, Kwapil, Walsh, & Myin-Germeys, 2014). Because the individual items vary significantly in their means and to reduce measurement error, the responses were combined into a single response score using a graded response model, a type of item response theory model. The graded response model is also inherently able to deal with the planned missingness in the scale. The scale scores are approximately centered at 0 with unit variance ($M = 0.06$, $SD = 0.90$).

To measure collective self-esteem, I used items adapted from the *public* and *private* subscales of Luhtanen and Crocker's (1992) collective self-esteem scale as well as related items from the Multidimensional Inventory of Black Identity which expands upon the measure of collective self-esteem (Sellers et al., 1998). In addition to adjusting items to refer to the political identity, the wording makes clear that the items refer to how one feels *right now* or *today*, rather than in general. Like the measure of identity strength, a random selection of 4 items is used in each daily survey to reduce the questionnaire length. A graded response model is also used to generate scores, which are likewise on an arbitrary scale with an approximate mean of 0 and unit variance (M = 0.06, SD = 0.84).

Each day, respondents were asked to estimate the amount of time they spent talking with others about news or politics the previous day. The total average was just over 20 minutes (M = 21.54, SD = 30.70, Median = 11.63), the majority of which was with co-partisans (M = 12.75, SD = 20.49, Median = 5.15), with out-partisans (M = 5.07, SD = 8.78, Median = 1.74) and people with either no partisanship or support a minor party (M = 3.72, SD = 7.68, Median = 0.71) rounding out the time spent in discussion. In models, a variable for the total amount of time spent in discussion is used to represent the general effect of discussion, separate from in-party discussion.. Respondents were likewise asked to estimate the amount of time spent engaging with news or political media in the previous days. The total average was slightly less than half an hour (M = 28.31, SD = 40.01, Median = 16.43). Respondents spent the most time with sources that support one's own party (M = 14.30, SD = 25.25, Median = 6.00) followed by time using non-partisan sources (M = 8.83, SD = 16.59, Median = 2.83) and sources supporting the other party (M = 5.17, SD = 11.55, Median = 1.27). Models include a variable for the total amount of

time spent using media to establish a baseline effect of media use aside from specifically in-party media sources.

As implied by the descriptive statistics, these measures are strongly right/positive skewed. To make the measures more amenable to analysis, two methods were used. The first was Winsorizing the measures (Tukey & McLaughlin, 1963; Wilcox, 2005), which means replacing extreme values with a less extreme one. In this case, any value beyond the 99.5th percentile of responses — 300 minutes (5 hours) for both measures — was replaced with the value of the 99.5th percentile. Upon close inspection, some of these values were likely entered in error, but rather than delete them (a procedure known as trimming) I take the more conservative route. The reported descriptive statistics are of the Winsorized data. Additionally, to attenuate the still-present right skew, I apply a natural log transformation to the media measures before analysis. Because many values are 0 and the log of 0 is undefined, the computation takes the form of $\log(x + 1)$ where x is the value being transformed. This has the useful property of retaining 0 as 0 since $\log(1) = 0$. The underlying logic of doing this transformation, besides the practical modeling concerns, is that (for instance) the substantive difference between 0 minutes of communication and 60 minutes of communication is greater than the difference between 60 minutes and 120 minutes.

Several other constructs were measured in the first survey that serve as controls in models to help address confounding for between-subjects comparisons. These include age (M = 21.36, SD = 3.39), race and ethnicity (86% white, 10% Black, 5% Asian, 6% Hispanic, 2% Middle Eastern/North African in a measure that allowed for multiple selections), and gender (74%

women). Respondents were asked 4 questions meant to tap aspects of ideology. These items assessed attitudes about the extent government should be involved in the provision of healthcare, whether the government should provide more or fewer services, whether the defense budget should be increased or decreased, and whether it is the government's responsibility to improve the socioeconomic position of racial and ethnic minorities. These items, measured on a scale from 1 to 7, were combined into a single measure that I will refer to as "issue alignment" and scaled such that higher values correspond with more ideological alignment with the respondent's preferred party (M = 4.91, SD = 1.22). In other words, a Republican respondent who gave the strongest "limited government" responses on each item would have a score of 7 while a Democrat giving the same answers would have a score of 1. Respondents were also asked if they had ever voted in an election in the United States; 56% said they had.

Some additional measures were not used as controls, largely because they would be too redundant with other measures of interest, but are mentioned here to help better describe the sample. In addition to the issue-based measure of ideology, respondents were also asked which ideological label described them on a scale from "very liberal" (1) to "very conservative" (7) with "moderate/middle of the road" as midpoint (M = 3.37, SD = 1.41). 56% chose some variation of liberal, 19% conservative, and 25% moderate. Democrats were asked about their preference in the then-ongoing presidential primary elections. 41% chose Bernie Sanders, 18% Joe Biden, 17% Elizabeth Warren, 9% Andrew Yang, 7% Pete Buttigieg, 7% said they would not support any of the candidates, and no respondents offered a candidate not listed (Mike Bloomberg's campaign launched after the study received ethics approval but before it went into the field).

To get a more detailed sense of communication behaviors, the initial survey asked about *typical* behaviors. Asked about how many different people they had talked about news or politics with in the past month, the typical respondent said about 5 of them (M = 7.18, SD = 8.99, Median = 5.00). Of those people, respondents estimated that 62% were supporters of the same party, 26% were supporters of the other major party, and 12% did not support a major party. Of note is that in the daily measures reported earlier, 59% of time talking politics was reported to be with co-partisans, 24% of time with out-partisans, and 17% with non-partisans/minor party supporters. The percentage of discussion partners from the same party was also correlated with the percentage of time spent talking to people supporting the same party at the individual level, $r = 0.41$. When including only those who reported discussion in both the introductory survey and in the daily surveys, the correlation increases to 0.52. Asked how many days in a typical week they talk about news or politics, respondents said on average just under 3 days (M = 2.69, SD = 1.98). This is rather close to the observed proportion of days in which respondents recall having discussions in the daily measures (on average 43% of days, i.e. 3.07 days per week). Figure 8 shows the relationship between these measures; there is plenty of error at the individual level, but overall the self-reported typical amount of discussion is clearly associated with day-to-day recollections of discussion. Although not unassailable evidence of the two measures' validity,

this consistency suggests the two are likely measuring the same construct.

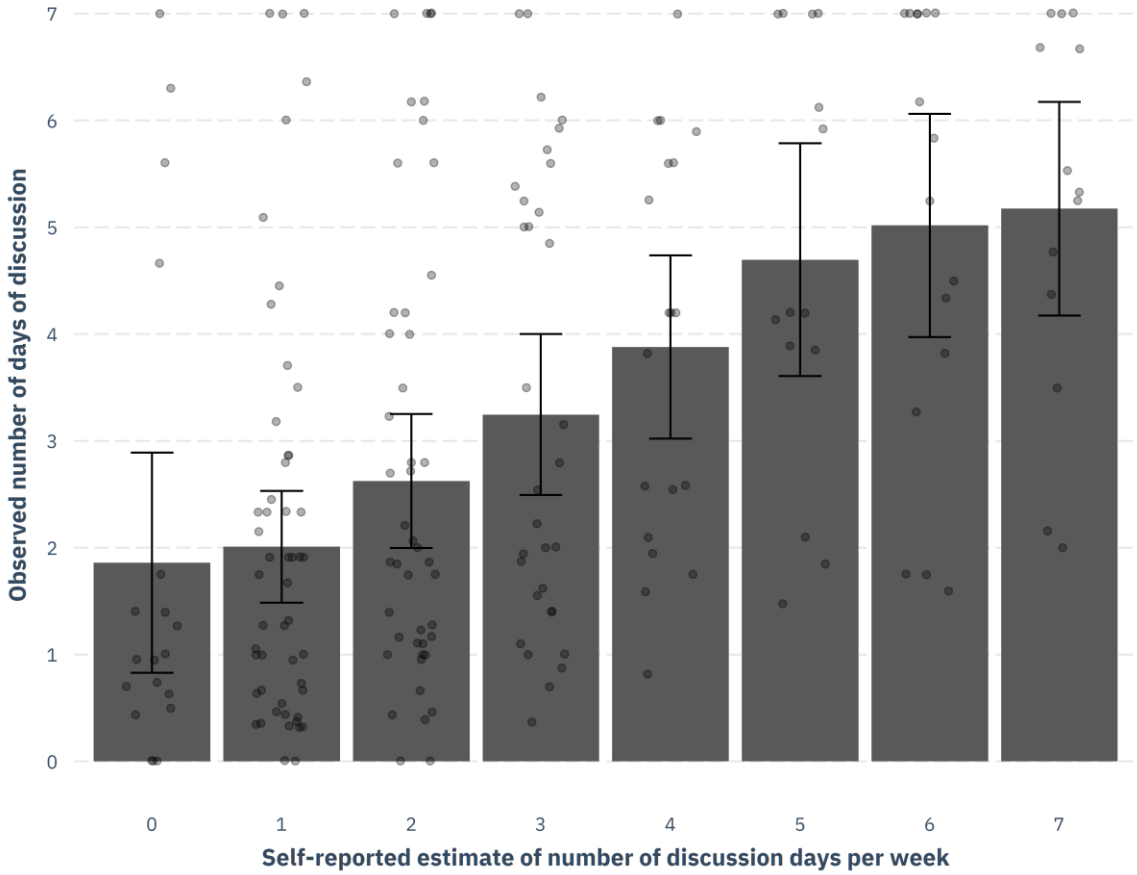


Figure 8: Comparison of Discussion Measures

To get an additional measure of media use in the introductory survey, respondents were given a list of 71 news sources, for which they were instructed to check a box besides each that they use regularly, defined as at least once in the past month. The prompt specified that sources reached via social media should be included. These sources included television, newspaper, and online-only outlets. This is a variation of the program list measure promoted as a best practice

for self-reported media measures (Dilliplane et al., 2013) and is quantified as the number of sources checked (M = 7.01, SD = 5.51, Median = 6.00). Although the number of sources is often taken as a proxy for amount of media use, it is not nearly as amenable to comparison to the daily measure used in this study as the measures of discussion. The correlation between the program list and the reported amount of time spent using political media per day was 0.24, which is statistically significant ($p < .001$) but rather low for measures that in scientific practice would be treated as interchangeable. It is no secret that self-reported measures of media are prone to measurement error (Guess, 2015; Prior, 2013b; Scharkow, 2019), but in this case it is not obvious whether the reason for the relatively low correlation between these measures is because of flaws in one, both, or neither of them. Although these two kinds of measures are typically treated equivalently in applied research, and one would expect a stronger correlation if they were measured without error, they are not on their face measurements of the same thing. Intense use of a small number of sources could result in a large amount of time spent exposed to media without a high number of sources used, for instance.

Variation in all measures can be partitioned into within-subject variance (σ_w^2) and between-subject variance (σ_b^2), which is relevant to the earlier chapter's discussion of the extent to which constructs are more like a trait (all true variance is between subjects) or like a state (all true variance is within subjects). Note that measurement error is generally considered to be a part of the within-subject variance. One way to express this partitioning of within- and between-subject variance is the intraclass correlation (ICC), defined for these purposes as $\frac{\sigma_b^2}{\sigma_b^2+\sigma_w^2}$, the portion of all variance that is explained by the subject (Stanish & Taylor, 1983). Another way to

symbolize this is in the multilevel modeling framework. Where Y_{it} is the variable of interest for subject i at time t, we can represent this as a multilevel linear regression with a random intercept:

$$Y_{it} = \alpha_i + \epsilon_{it}$$

Where α_i is the subject-specific mean for subject i and ϵ_{it} is error, with a mean of 0. "Error" here simply means all variation that is not explained by the subject's own mean; it encompasses both true within-subject variance and measurement error. In this framework, the ICC is equivalently represented as $\frac{\sigma_\alpha^2}{\sigma_\alpha^2+\sigma_\epsilon^2}$, so the variance in subject means α is equivalent to σ_b^2 and the error variance is equivalent to σ_w^2. It is not possible without invoking assumptions from measurement theory to further partition σ_w^2 into true variance and measurement error — it in fact requires assumptions from measurement theory to even presuppose that all measurement error is absorbed into σ_w^2. Instead, I will just describe the between-within distinction and readers can consider that it is likely that some portion of the within-subject variance is due to measurement error.

Finally, when reporting ICCs with data such as these, one must decide whether to perform the calculations using only the observed data or by using the multiply imputed data instead. Research indicates that ICC is likely to be overestimated when ignoring the missing observations. Specifically, σ_w^2 can only be underestimated when data are missing, never overestimated (Newman & Sin, 2009), which means any non-zero relationship between the variable and response will bias σ_w^2 downward. Furthermore, evidence suggests non-responders

give more variable responses than responders (Rogelberg et al., 2003; Rogelberg, Luong, Sederburg, & Cristol, 2000). Of course, the multiply imputed data can give inaccurate estimates of ICC as well. The most relevant prior work would indicate that a well-specified imputation model can generate unbiased ICC estimates and mis-specified models tend to overestimate ICC (Lüdtke, Robitzsch, & Grund, 2017). I will report both and with this information, readers can treat the complete-data ICC as an upper bound. The multiply imputed estimate cannot be interpreted as a lower bound, however, since it could be wrong in either direction.

For identity strength, the ICC using only complete data is 0.85, meaning 15% of variance is within subjects. With the imputed data, the ICC is instead 0.66, implying 34% of variance is within subjects. For collective self-esteem, as theoretically expected the ICC estimates are a lower 0.77 and 0.60 for the complete and imputed data, respectively. The ICC estimates for the (log-transformed) amount of media consumption are 0.54 and 0.46. Similarly, the ICCs for total amount of discussion are 0.50 and 0.43. Given the prior research, it is unsurprising that the ICC is lower for the communication variables simply because they are reputed to be very high in measurement error. Of course, as discussed, the actual amount of communication should also be expected to change more within people on a day to day basis than their sense of political identity and self-esteem would. Within-subject variability also encompasses both time-dependent and net variability as described in previous chapters, but does not distinguish the two. Overall, it is clear that these variables have sufficient within-subject variation to make within-subject analyses worthwhile and these analyses are required in order to establish the extent to which any of that variation reflects development, net variability, and time-structured variability.

Analysis Plan

For analysis, I used the multilevel AR(1) model described previously for each of the time-varying outcomes (Hamaker, Asparouhov, Brose, Schmiedek, & Muthén, 2018; Hedeker et al., 2012). This is conceptually similar to a cross-lagged panel model, but it disaggregates the within-person from between-person effects as is done in the approach econometricians call fixed effects modeling. Such models can be estimated using Markov Chain Monte Carlo (MCMC) simulation, which allows for arbitrarily complex regression models in a computationally intensive, Bayesian framework. Multiple models were estimated: one for each of identity strength, collective self-esteem, identity-affirming media use, and identity-affirming discussion. Each variable will predict the other, with lagged values of identity strength and collective self-esteem predicting the communication variables while the non-lagged communication variables may be the most appropriate since the measures refer to the previous day[16].

There are two ways I have discussed operationalizing variability in previous chapters. One has to do with autocorrelation of deviations from the mean from one occasion to the next and the other regards the magnitude of those deviations. The best way to probe the effect on the

[16] Because effect estimates can be badly biased by misspecified lags (Leszczensky & Wolbring, 2019), I explored multiple specifications to test their sensitivity and gain more clarity on the true timing of effects. Although model comparison alone cannot adjudicate the timing of effects and the appropriateness of any given specification, models with a greater number of lags indicated the specification described captured effects in a parsimonious way.

autocorrelation, for which values close to 0 reflect more stability, is to include interaction terms between key predictors and the autocorrelation term. For instance, consider a model with identity strength as the outcome variable. An interaction between the lagged value of identity strength and identity-affirming communication will estimate the extent to which identity-affirming communication affects the autocorrelation of identity strength. This is, in my framework, a primarily between-subject phenomenon — that is, what affects this kind of stability is not where a person is relative to his or her norms, but the actual level of the variables. For instance, my hypothesis is that a high level of affirming communication should contribute to stability. This does not mean I expect added stability when a person who engages in zero communication decides one day to spend a few minutes watching the news. Rather, I expect someone who usually spends a good deal of time engaging in communication to demonstrate more stability regardless of their day-to-day changes. Hypotheses regarding development, or longer-lasting changes in the mean levels of an outcome, are tested with the main effects in these models and are conceptually equivalent to the typical kind of media effects. A positive main effect does not literally mean that the effect is long lasting, but empirically shows that changes in the predictors are associated with changes in the outcome beyond what can be explained by the outcome's mean level and overall trend. These effects are investigated solely by focusing on the within-person effects, which can be interpreted causally under assumptions that are weaker than for between-person effects and in cross-sectional designs. Models also include the time variable, since accurate estimates for variability (autocorrelation and residual variance) depend on the outcome variables being stationary conditional on the predictors. This means, in effect, that the models are a variant of the more familiar growth curve models.

MCMC estimation does not provide the analyst with the usual tools of statistical inference like test statistics, *p* values, and the like, but approximations are available. Simple methods of doing so include 95% credibility intervals, which are a Bayesian counterpart to confidence intervals that are interpreted as having a 95% probability of containing the true value. Likewise, posterior probabilities, which are Bayesian counterparts for *p* values, have the intuitive interpretation of being the probability the parameter is greater/less than 0 (Makowski, Ben-Shachar, Chen, & Lüdecke, 2019). The key to interpretation of posterior probabilities is that a 95% posterior probability is equivalent in interpretation to a one-sided *p* value of .05, although as a general rule Bayesian analyses tend to be more conservative. The existence of an effect will be assessed primarily using the posterior probability, using the 95% credibility interval as an additional indicator of uncertainty.

The models are estimated using the R package "brms" (Bürkner, 2018), which interfaces with the MCMC estimation software "Stan" (Carpenter et al., 2017) using the No-U-Turn sampler. Stan has a number of advantages to the end user compared to BUGS, which was used by Jongerling, Hamaker, and collaborators (2015; 2012), but produces substantively identical results while avoiding efficiency and convergence problems. It is important when doing Bayesian inference to report the prior distributions used in estimation (Depaoli & van de Schoot, 2017), which reflect prior knowledge (or lack thereof) about the values to be estimated. Conventionally, one centers the prior distributions on parameters like regression coefficients to 0 in order to state a prior belief of a null effect. This places the burden on the data to demonstrate non-zero effects. For the regression estimates and the random effect deviations, I specified a normal prior distribution centered at 0 with a standard deviation of 1. This is a "weakly

informative" prior (Gelman, Jakulin, Pittau, & Su, 2008) in that it places a gentle constraint on the estimates to stay within plausible values but allows the model to estimate more extreme parameters if they are clearly justified by the data. For estimation, 3 MCMC chains of 4000 iterations each are used, of which half are considered warmups and discarded. This is done on each of the imputed datasets, then the chains are combined to create a single model with 75 chains for each parameter (3 chains for each of 25 datasets). Computation was performed on the Unity high-performance computing cluster managed by The Ohio State University College of Arts and Sciences.

Chapter 9. Results

I begin with an overall summary of results before discussing them in detail for each dependent variable. To reiterate, 3 types of effects are investigated: development, time-structured intraindividual variability, and net intraindividual variability. Table 1 summarizes results for hypotheses and Table 2 summarizes results relevant to research questions. One finding that is consistent across all dependent variables in this study is that there is very limited evidence of any effects on time-structured intraindividual variability. Although the reasons for the lack of effects in that domain will be more fully explored in the discussion, 3 of the 4 dependent variables did not have significant autocorrelation across observations, which as previously discussed is generally interpreted to be evidence of a certain kind of stability. On the other hand, there are several effects of interest on net intraindividual variability. The variable whose variability is of most interest theoretically is identity strength. Results show that although discussion in general increases variability of identity strength, discussion with supporters of the same party reduces variability, as hypothesized. In the case of media, there is no general effect across all types of media, but in-party media in particular show some evidence of reducing the variability of identity strength in line with expectations. Consistent with expectation, identity strength predicts less net variability of collective self-esteem. On the other hand, contrary to predictions, identity strength was associated with greater variability of in-party media and discussion. As a note on terminology, although I have discussed "identity-affirming" communication when discussing theory, I will refer to the measured variables as "in-party" discussion and media to more clearly convey exactly what was measured.

Some effects on the mean levels of the dependent variables were observed, which I have referred to as development and are the usual sort of communication effect. For these, the sole focus is on within-person effects. As expected, increases in collective self-esteem appear to cause increases in identity strength. The other causal direction for this relationship was also observed, meaning that increases in strength appear to cause increases in collective self-esteem. Discussion with supporters of the same party as well as use of in-party media sources also had direct, positive effects on both collective self-esteem and identity strength, as predicted. Addressing one of the research questions, media in general, controlling for the level of in-party media, appear to cause reductions in identity strength and collective self-esteem, although the latter effect has weaker evidence. For models with in-party communication as the dependent variable, neither identity strength nor collective self-esteem appear to cause changes in in-party communication. Although in-party media appears to promote increases in in-party discussion, the reverse causal direction is not supported by the results. That being said, both models with communication as dependent variable have some important statistical caveats to be discussed later.

Table 1: Summary of Support for Hypotheses

Hypothesis	Type of Effect	Summary of Results
H1: In-party communication will decrease variability of identity strength.	Net variability	**Talk**: Supported **Media**: Weak support
H1: In-party communication will decrease variability of identity strength.	Time-structured variability	**Talk**: Not supported **Media**: Not supported
H2: In-party communication will increase identity strength.	Development	**Talk**: Supported **Media**: Supported
H3: In-party communication will increase collective self-esteem.	Development	**Talk**: Supported **Media**: Supported
H4: Collective self-esteem will be positively related to identity strength.	Development	Supported
H5: Identity strength willdecrease variability of in-party communication.	Net variability	**Talk**: Not supported **Media**: Not supported
H5: Identity strength willdecrease variability of in-party communication.	Time-structured variability	**Talk**: Not supported **Media**: Not supported
H6: Increases in identity strength will increase in-party communication.	Development	**Talk**: Weak support **Media**: Not supported

Table 2: Summary of Results Pertaining to Research Questions

Research Question	Type of Effect	Summary of Results
RQ1: Does the extent to which in-party communication promotes variability and development of identity strength depend on the level of identity strength?	Net variability	**Talk**: No **Media**: No
RQ1: Does the extent to which in-party communication promotes variability and development of identity strength depend on the level of identity strength?	Development	**Talk**: No **Media**: No
RQ2: Does collective self-esteem influence in-party communication?	Net variability	**Talk**: No **Media**: No
RQ2: Does collective self-esteem influence in-party communication?	Development	**Talk**: No **Media**: No
RQ3: Does identity strength influence collective self-esteem?	Development	Yes, increases
RQ3: Does identity strength influence collective self-esteem?	Net variability	Yes, people with high identity strength have lower variability of collective self-esteem
RQ4: How do the effects of non-in-party communication compare with those of in-party communication?	Development	**Talk**: No effect on any dependent variables **Media**: Decreases identity strength and collective self-esteem
RQ4: How do the effects of non-in-party communication compare with those of in-party communication?	Net variability	**Talk**: Greatly increases variability of identity strength and collective self-esteem; increases variability of in-party talk; reduces variability of in-party media **Media**: No effect on identity strength; decreases variability of self-esteem; increases variability of in-party media; decreases variability of in-party talk

Identity Strength as Outcome

The model predicting identity strength is perhaps the most important, theoretically speaking. I have suggested that this is the variable most amenable to study because it should be variable enough to detect meaningful changes. On a statistical level, it is both well-measured and normally distributed, making model specification and interpretation simpler. Before discussing the hypotheses and research questions, I first want to draw attention to some parameters of the model to better explain how some of the tests are implemented and explore some more basic theoretical issues. The distinctive feature of multilevel models is that coefficients can, if the analyst allows, vary by level (in this case, the levels are respondents). These varying coefficients are generally called "random effects." By convention, the intercept of multilevel models is always modeled as a random effect and each participant's intercept can be interpreted as an (conditional) estimate of their mean of the dependent variable, net of the predictor variables. As expected, participant intercepts for identity strength vary even after accounting for control variables, with an estimated standard deviation of 0.42 (95% CI [0.37, 0.48], posterior probability >99.9%), consistent with a conceptualization of strength as being at least partly trait-like. Another random effect in the model is for time, allowing each participant to have their own underlying growth curve over the course of the study (SD = 0.02, 95% CI [0.01, 0.02], posterior probability >99.9%).

In the estimation of random effects, one does not typically just allow each random effect to vary but also allows them to correlate with one another. Although these correlations are often

considered substantively uninteresting, for these purposes there are theoretically relevant random effect correlations. I have previously argued that one explanation for why the mutual reinforcement of identity and media use does not inevitably lead to extremity is that there may be an inherent pull toward moderation for those with high identity strength. In this model, each participant has their own estimate of both identity strength (the random intercept) and of their time trend (the random growth curve) after accounting for other variables. The correlation between these is -0.300 (95% CI [-0.467, -0.112], posterior probability 99.3%), suggesting that people with high average identity strength do have a tendency to lose strength over time *if* other factors like communication are assumed to be 0. This is unlikely to be a ceiling effect because the measure is designed to prevent ceiling effects. It is unlikely to be a regression to the mean effect because the correlation is between the time trend and the overall mean, not the first measure of strength. Note also that this does not necessarily mean that people with relatively high identity strength actually did tend to see an over-time decrease in identity strength, because this estimate is only net of the other variables, many of which affect the level of identity strength.

Now, I turn to results that bear directly on hypotheses and research questions. RQ1 asks whether the effect of in-party communication on identity strength is itself moderated by identity strength. In other words, RQ1 asks whether people with higher or lower identity strength have different susceptibility to influence from communication. To test the dependence of effects on identity strength, the normal strategy — including an interaction term — is not viable, since it is not computationally feasible to include the observed mean level of identity strength as a regressor when the random intercept is nearly equivalent to the observed mean. Instead, I estimated the effects of in-party discussion and media use as random effects, which could then

correlate with the random intercept in the same way I just discussed for the growth curves. There was little indication for the effects of in-party media depending on the level of identity strength (r = -0.347, 95% CI [-0.769, 0.169], posterior probability 88.5%) and no evidence for identity-affirming discussion (r = -0.110, 95% CI [-0.518, 0.317], posterior probability 68.2%). The estimated correlations are consistent with the notion that effects of in-party discussion are attenuated for those stronger identities, which is another potential explanation for why people do not tend to become extreme. That being said, these associations do not permit any inferences to that effect.

Contrary to the prediction of H1, no predictor appeared to affect the time-dependent variability of identity strength, which is operationalized as interaction terms with the lagged value of identity strength (also included in Table 3). This could partly be explained by the fact that the autocorrelation term itself was indistinct from 0 (B = 0.067, 95% CI [-0.021, 0.154], posterior probability 93.4%), meaning that deviations from one's typical level of identity strength on one day do not tend to be carried over to the next day or get overcompensated and cause lower-than-normal levels on the next day. This is generally interpreted as evidence of stability in the sense that a person self-regulates well. This finding may raise the question: Do participants, as a whole, just not exhibit this kind of variability? The evidence suggests that some do. The model includes a random effect for the autocorrelation coefficient and results indicate that participants differ from one another significantly in terms of their own autocorrelation from one occasion to the next. The individual-specific autocorrelation has standard deviation of 0.14 (95% CI [0.08, 0.19], posterior probability >99.9%). This is evidence that some participants do have non-zero autocorrelation coefficients, but across participants they are not consistently positive or negative.

Substantively, it is most likely that many self-regulate well (meaning their autocorrelation is 0), but some non-trivial proportion do not. What the data do not show is what the underlying factors associated with better or worse self-regulation are. Another clue that communication and/or collective self-esteem may play a role: If these interaction terms are omitted from the model, then there is significant, positive autocorrelation. This means it appears that these variables do affect time-dependent variability *in aggregate.*

H2 predicts that in-party communication will lead to increases in identity strength. The model detects a positive effect of in-party discussion (B = 0.033, 95% CI [0.005, 0.060], posterior probability 99.2%), indicating that increases in discussion with supporters of the same party is associated with greater identity strength on the next day. The estimate for in-party media is similar (B = 0.039, 95% CI [0.015, 0.062], posterior probability 99.9%). H2 is supported by the analysis. As predicted by H4, increases in collective self-esteem are associated with greater identity strength the following day (B = 0.041, 95% CI [0.009, 0.073], posterior probability 99.6%). Addressing RQ4 about general effects of communication besides in-party sources, the model shows a negative effect of media (B = -0.024, 95% CI [-0.040, -0.008], posterior probability 99.9%). Discussion, however, did not have any apparent effect (B = -0.009, 95% CI [-0.029, 0.010], posterior probability 81.5%). As within-subject effects, these can be reasonably interpreted as causal. Full results are in Table 3.

Table 3: Regression Results for Effects on Mean Level of Identity Strength

	Estimate	95% CI	Post. Prob.
Within-subject effects			
Strength (t – 1); autocorrelation	0.067	[-0.021, 0.154]	93.4%
Esteem (t – 1)	0.041	[0.009, 0.073]	99.6%
In-party talk	0.033	[0.005, 0.060]	99.2%
In-party media	0.039	[0.015, 0.062]	99.9%
Total talk	-0.009	[-0.029, 0.010]	81.5%
Total media	-0.024	[-0.040, -0.008]	99.9%
Effects on time-dependent variability			
Esteem	0.037	[-0.020, 0.094]	89.9%
In-party talk	-0.028	[-0.162, 0.111]	64.8%
In-party media	-0.002	[-0.102, 0.098]	51.8%
Total talk	0.074	[-0.031, 0.181]	91.3%
Total media	-0.032	[-0.116, 0.052]	77.3%
Between-subject effects			
Esteem	0.835	[0.739, 0.929]	>99.9%
In-party talk	0.108	[-0.090, 0.307]	85.8%
In-party media	0.068	[-0.075, 0.212]	82.4%
Non-partisan media	0.036	[-0.130, 0.202]	66.5%
Out-party talk	-0.067	[-0.200, 0.064]	84.0%
Republican	0.156	[0.007, 0.309]	97.9%
Gender (woman)	-0.009	[-0.149, 0.137]	54.9%
Issue alignment	0.005	[-0.058, 0.067]	56.2%
Race/ethnicity (White)	-0.247	[-0.465, -0.027]	98.6%
Race/ethnicity (Hispanic)	-0.311	[-0.570, -0.046]	99.0%
Race/ethnicity (Black)	-0.020	[-0.272, 0.230]	56.2%
Age	-0.010	[-0.028, 0.008]	85.4%
Voted before	0.174	[0.046, 0.300]	99.6%
Survey number	0.001	[-0.002, 0.005]	74.0%
Intercept	0.205	[-0.380, 0.776]	75.6%

Note: "Post. Prob." refers to posterior probability, which readers can interpret approximately as 1 – p.

Turning to net intraindividual variability, the other part of H1, there are several effects of note (see Table 4 for all details). At the between-person level, which is where the focus is for net

variability, in-party discussion is associated with decreased variability (B = -0.251, 95% CI [-0.455, -0.058], posterior probability 99.4%). The estimate for in-party media is also negative but statistically not clearly distinct from 0 (B = -0.105, 95% CI [-0.254, 0.034], posterior probability 92.5%). Relevant to RQ4, the single largest predictor of net variability is the total amount of discussion (B = 0.448, 95% CI [0.289, 0.605], posterior probability >99.9%). This swamps the collective effects of nearly all other variables in the model.

Table 4: Regression Results for Effects on Residual Variance of Identity Strength

	Estimate	95% CI	Post. Prob.
Within-subject effects			
Strength	-0.028	[-0.119, 0.081]	74.0%
Esteem ($t - 1$)	-0.075	[-0.130, -0.022]	99.8%
In-party talk	-0.030	[-0.077, 0.015]	90.2%
In-party media	0.109	[0.065, 0.152]	>99.9%
Total talk	0.066	[0.033, 0.099]	>99.9%
Total media	0.010	[-0.045, 0.068]	62.4%
Between-subject effects			
Intercept	-1.154	[-1.693, -0.622]	>99.9%
Esteem	0.114	[0.001, 0.225]	97.5%
In-party talk	-0.251	[-0.455, -0.058]	99.4%
In-party media	-0.105	[-0.254, 0.034]	92.5%
Total talk	0.448	[0.289, 0.605]	>99.9%
Total media	-0.035	[-0.164, 0.102]	70.2%
Republican	-0.051	[-0.197, 0.090]	76.0%
Gender (woman)	0.026	[-0.112, 0.161]	64.3%
Issue alignment	-0.016	[-0.074, 0.045]	69.8%
Race/ethnicity (White)	-0.125	[-0.333, 0.084]	88.0%
Race/ethnicity (Hispanic)	0.077	[-0.171, 0.329]	72.5%
Race/ethnicity (Black)	-0.071	[-0.304, 0.164]	72.8%
Age	0.004	[-0.012, 0.022]	69.2%

	Estimate	95% CI	Post. Prob.
Voted before	-0.122	[-0.256, 0.002]	97.1%

Note: "Post. Prob." refers to posterior probability, which readers can interpret approximately as 1 – *p*.

Collective Self-esteem as Outcome

For the sake of parsimony, I will not discuss the random effects parameters of the model with collective self-esteem except to say the random intercept variance between subjects is clearly significant for both the main and residual variance parts of the model. Unlike the dependent variables in other models, collective self-esteem does show some evidence of autocorrelation (B = 0.083, 95% CI [-0.004, 0.169], posterior probability 97.0%), meaning a change on one day is expected to partly carry over into the next. This is consistent with collective self-esteem being more state-like than identity strength, although it is also clear that collective self-esteem has a trait-like quality as well. Addressing RQ3, just as collective self-esteem affected identity strength, this model shows support for the reverse causal path as well (B = 0.054, 95% CI [0.017, 0.093], posterior probability 99.8%). Supporting H3, in-party discussion (B = 0.032, 95% CI [0.005, 0.058], posterior probability 99.1%) and in-party media (B = 0.025, 95% CI [0.002, 0.049], posterior probability 98.5%) both appear to increase collective self-esteem. Total media use is weakly associated with decreased collective self-esteem (B = -0.013, 95% CI [-0.031, 0.000], posterior probability 97.3%), relevant to one of the research questions and similar to its effect on identity strength. The estimated effect for all types of discussion was similar to that for media, but not clearly distinct from 0 (B = -0.016, 95% CI [-0.032, 0.008], posterior probability 88.5%). As for identity strength, no variable clearly had a significant effect on time-dependent variability, but there is some indication that identity strength may increase

autocorrelation (interaction B = 0.047, 95% CI [-0.007, 0.102], posterior probability 95.4%). This was not an effect that was hypothesized so I will not dwell on it any further. Full details for this part of the model are included in Table 6. Because no specific hypotheses or research questions were enumerated regarding the variability of collective self-esteem, I will not review them in the text. Full details of the net variability part of the model are in Table 7.

Table 5: Regression Results for Effects on Mean Level of Collective Self-esteem

	Estimate	95% CI	Post. Prob.
Within-subject effects			
Esteem (t – 1); autocorrelation	0.083	[-0.004, 0.169]	97.0%
Strength (t – 1)	0.054	[0.017, 0.093]	99.8%
In-party talk	0.032	[0.005, 0.058]	99.1%
In-party media	0.025	[0.002, 0.049]	98.5%
Total talk	-0.013	[-0.032, 0.008]	88.5%
Total media	-0.016	[-0.031, 0.000]	97.3%
Effects on time-dependent variability			
Strength	0.047	[-0.007, 0.102]	95.4%
In-party talk	-0.065	[-0.204, 0.069]	82.2%
In-party media	-0.019	[-0.123, 0.086]	64.4%
Total talk	0.050	[-0.053, 0.154]	82.8%
Total media	-0.007	[-0.090, 0.078]	56.4%
Between-subject effects			
Intercept	-0.381	[-0.922, 0.148]	91.9%
Strength	0.736	[0.649, 0.819]	>99.9%
In-party talk	0.084	[-0.106, 0.272]	80.9%
In-party media	-0.015	[-0.150, 0.118]	58.8%
Non-partisan media	-0.073	[-0.225, 0.074]	83.3%
Out-party talk	0.042	[-0.080, 0.162]	75.3%
Republican	-0.071	[-0.209, 0.070]	84.0%
Gender (woman)	0.177	[0.045, 0.304]	99.6%
Issue alignment	0.038	[-0.021, 0.096]	90.1%
Race/ethnicity (White)	0.085	[-0.120, 0.290]	79.3%
Race/ethnicity (Hispanic)	0.111	[-0.148, 0.365]	80.2%
Race/ethnicity (Black)	-0.027	[-0.256, 0.201]	59.1%
Age	0.002	[-0.015, 0.019]	58.5%
Voted before	-0.092	[-0.211, 0.029]	93.2%

Note: “Post. Prob.” refers to posterior probability, which readers can interpret approximately as 1 – p.

Table 6: Regression Results for Effects on Residual Variance of Collective Self-esteem

	Estimate	95% CI	Post. Prob.
Within-subject effects			
Strength (t – 1)	-0.006	[-0.057, 0.049]	59.6%
In-party talk	-0.076	[-0.121, -0.028]	99.9%
In-party media	-0.016	[-0.059, 0.028]	76.2%
Total talk	0.101	[0.066, 0.138]	>99.9%
Total media	0.064	[0.028, 0.099]	>99.9%
Between-subject effects			
Intercept	-1.355	[-1.905, -0.821]	>99.9%
Strength	-0.042	[-0.123, 0.045]	83.8%
In-party talk	-0.284	[-0.475, -0.094]	99.8%
In-party media	0.038	[-0.104, 0.175]	70.4%
Total talk	0.460	[0.310, 0.616]	>99.9%
Total media	-0.144	[-0.267, -0.017]	98.7%
Republican	0.035	[-0.106, 0.176]	68.9%
Gender (woman)	-0.015	[-0.146, 0.120]	59.0%
Issue alignment	0.027	[-0.031, 0.087]	81.7%
Race/ethnicity (White)	-0.076	[-0.293, 0.137]	75.8%
Race/ethnicity (Hispanic)	0.145	[-0.103, 0.394]	87.3%
Race/ethnicity (Black)	-0.032	[-0.270, 0.207]	60.8%
Age	0.012	[-0.005, 0.028]	91.2%
Voted before	-0.073	[-0.196, 0.047]	88.1%

Note: "Post. Prob." refers to posterior probability, which readers can interpret approximately as 1 – *p*.

In-party Discussion as Outcome

For both models with a communication variable as dependent variable, there are a few details regarding model specification worth noting. First, models could not converge when a random intercept was included in the residual variance part of the model, so it is omitted. This

increases the risk of omitted variables bias in that part of those models. Next, for in-party discussion (media), the total amount of discussion (media) on that day is included as a control variable. This has an effect on interpretation such that effects should be interpreted as about the partisan composition of the day's discussion (media) moreso than the amount of it. There are substantive arguments for and against this specification and it was chosen partly to aid in computation — the residuals of the model without this control added are extremely non-normal and not very amenable to this kind of analysis. Overall, these statistical difficulties mean these models should be interpreted more cautiously. There are statistical models that better capture the true distribution of the communication dependent variables, but they do not have equivalent variance parameters with which hypotheses about net variability can be tested.

H6 predicted that increases in identity strength would promote more in-party discussion. The model provides weak evidence of an effect of identity strength (B = 0.049, 95% CI [-0.012, 0.110], posterior probability 93.9%). RQ2 asked whether collective self-esteem influenced in-party discussion. Increases in collective self-esteem were not associated with subsequent increases in in-party discussion. For main effects on in-party discussion, there are two of note (full details are in Table 8). Of note but not related to any a priori hypotheses, increases in in-party media are associated with more in-party discussion on the next day (B = 0.064, 95% CI [0.014, 0.114], posterior probability 99.5%), suggesting day-to-day variations in these modalities are not independent. There was one effect on time-dependent variability observed: When identity-affirming media and identity-affirming discussion increase at the same time, the increase in discussion becomes more likely to carry over into the subsequent day (i.e., in-party media induces autocorrelation in in-party discussion). H5 predicted identity strength would decrease net

variability of in-party discussion, but there was no main effect observed on the between-subjects level where the focus is and in fact the estimate is slightly positive. Full details for net variability are included in Table 9.

Table 7: Regression Results for Effects on Mean Level of In-party Discussion

	Estimate	95% CI	Post. Prob.
Within-subject effects			
In-party talk (t – 1); autocorrelation	-0.041	[-0.148, 0.065]	77.7%
Strength (t – 1)	0.049	[-0.012, 0.110]	93.9%
Esteem (t – 1)	0.000	[-0.055, 0.053]	50.3%
In-party media (t – 1)	0.064	[0.014, 0.114]	99.5%
Total media (t – 1)	0.009	[-0.021, 0.039]	72.7%
Total talk (t – 1)	0.011	[-0.023, 0.044]	74.0%
Effects on time-dependent variability			
Strength	0.004	[-0.082, 0.087]	53.4%
Esteem	0.044	[-0.043, 0.127]	84.7%
In-party talk	0.090	[0.017, 0.164]	99.2%
Total media	-0.026	[-0.095, 0.041]	76.9%
Total talk	0.021	[-0.045, 0.087]	72.8%
Between-subject effects			
Intercept	-0.231	[-0.594, 0.115]	90.0%
Strength	0.039	[-0.042, 0.120]	83.2%
Esteem	0.075	[-0.013, 0.161]	95.5%
In-party media	0.282	[0.197, 0.370]	>99.9%
Total media	-0.168	[-0.239, -0.096]	>99.9%
Total talk	0.494	[0.423, 0.564]	>99.9%
Republican	0.002	[-0.091, 0.091]	52.2%
Gender (woman)	-0.061	[-0.149, 0.025]	91.6%
Issue alignment	0.030	[-0.008, 0.067]	93.8%
Race/ethnicity (White)	0.070	[-0.062, 0.202]	85.1%
Race/ethnicity (Hispanic)	-0.063	[-0.242, 0.120]	75.4%
Race/ethnicity (Black)	0.030	[-0.122, 0.186]	64.4%
Age	0.006	[-0.005, 0.017]	86.9%
Voted before	0.024	[-0.056, 0.100]	72.6%
Survey number	-0.004	[-0.010, 0.001]	94.7%

Note: "Post. Prob." refers to posterior probability, which readers can interpret approximately as 1 – p.

Table 8: Regression Results for Effects on Residual Variance of In-party Discussion

	Estimate	95% CI	Post. Prob.
Within-subject effects			
Strength (t – 1)	0.031	[-0.034, 0.090]	83.2%
Esteem (t – 1)	-0.004	[-0.057, 0.048]	55.7%
In-party media (t – 1)	0.048	[0.013, 0.083]	99.7%
Total talk (t – 1)	-0.005	[-0.026, 0.015]	68.5%
Total media (t – 1)	0.016	[-0.011, 0.042]	87.6%
Between-subject effects			
Intercept	-1.366	[-1.692, -1.014]	>99.9%
Strength	0.062	[-0.006, 0.128]	96.0%
Esteem	0.032	[-0.033, 0.100]	82.6%
In-party media	0.122	[0.068, 0.176]	>99.9%
Total media	-0.030	[-0.085, 0.022]	86.8%
Total talk	0.205	[0.156, 0.258]	>99.9%
Republican	0.026	[-0.063, 0.110]	73.1%
Gender (woman)	-0.060	[-0.130, 0.007]	95.5%
Issue alignment	0.084	[0.050, 0.118]	>99.9%
Race/ethnicity (White)	0.244	[0.117, 0.387]	>99.9%
Race/ethnicity (Hispanic)	0.014	[-0.135, 0.160]	56.6%
Race/ethnicity (Black)	0.255	[0.128, 0.394]	>99.9%
Age	0.008	[0.000, 0.016]	98.0%
Voted before	-0.020	[-0.086, 0.043]	71.7%

Note: “Post. Prob.” refers to posterior probability, which readers can interpret approximately as 1 – p.

In-party Media as Outcome

The caveats and model specification details for the model with in-party discussion as dependent variable also apply to the model with identity-affirming media use as dependent variable. H6 predicted that increases in identity strength would promote greater use of in-party media, but no such effect was observed. Likewise, addressing RQ2, changes in collective self-esteem appeared unrelated to subsequent in-party media use. Overall, at the within-subject level, there were no robust predictors of changes in in-party media use. It is worth pointing out, then, that although in-party media use appears to promote in-party discussion, in-party discussion does not appear to promote in-party media use. This relationship was not the subject of any specific hypothesis, but it is nevertheless an interesting observation. Likewise, no autocorrelation over time was observed for identity-affirming media use. For time-dependent variability, quite consistent with all other models, no variable had any clear effect on the level of autocorrelation of identity-affirming media. Full results are in Table 10. H5 predicted identity strength would decrease net variability of in-party discussion, but similar to the in-party discussion model, strength is actually associated with a modest increase in variability. Other details of the net variability part of the model are included in Table 11.

Table 9: Regression Results for Effects on Mean Level of In-party Media

	Estimate	95% CI	Post. Prob.
Within-subject effects			
In-party media (t – 1); autocorrelation	0.014	[-0.087, 0.118]	61.0%
Strength (t – 1)	0.003	[-0.051, 0.056]	54.6%
Esteem (t – 1)	0.000	[-0.048, 0.048]	50.3%
In-party talk (t – 1)	0.036	[-0.013, 0.084]	92.1%
Total talk (t – 1)	0.011	[-0.024, 0.044]	72.9%
Total media (t – 1)	0.027	[-0.002, 0.056]	96.2%
Effects on time-dependent variability			
Strength	0.035	[-0.058, 0.126]	76.7%
Esteem	-0.061	[-0.155, 0.032]	89.8%
In-party talk	0.069	[-0.040, 0.178]	89.8%
Total talk	-0.082	[-0.180, 0.020]	94.7%
Total media	0.064	[-0.009, 0.137]	95.6%
Between-subject effects			
Intercept	0.052	[-0.430, 0.527]	58.7%
Strength	0.046	[-0.066, 0.158]	78.9%
Esteem	-0.008	[-0.131, 0.116]	55.2%
In-party talk	0.598	[0.440, 0.762]	>99.9%
Total media	0.483	[0.391, 0.577]	>99.9%
Total talk	-0.369	[-0.497, -0.239]	>99.9%
Republican	-0.058	[-0.178, 0.060]	83.4%
Gender (woman)	-0.048	[-0.168, 0.075]	78.2%
Issue alignment	0.002	[-0.047, 0.052]	52.7%
Race/ethnicity (White)	0.007	[-0.169, 0.181]	53.3%
Race/ethnicity (Hispanic)	0.031	[-0.206, 0.267]	60.6%
Race/ethnicity (Black)	-0.025	[-0.215, 0.173]	60.3%
Age	0.001	[-0.014, 0.016]	57.1%
Vote before	-0.115	[-0.220, -0.009]	98.5%
Survey number	-0.006	[-0.011, -0.001]	98.5%

Note: “Post. Prob.” refers to posterior probability, which readers can interpret approximately as 1 – *p*.

Table 10: Regression Results for Effects on Residual Variance of In-party Media

	Estimate	95% CI	Post. Prob.
Within-subject effects			
Strength (t – 1)	-0.025	[-0.101, 0.042]	75.3%
Esteem (t – 1)	-0.007	[-0.069, 0.054]	58.0%
In-party talk (t – 1)	0.030	[-0.011, 0.074]	92.0%
Total talk (t – 1)	0.028	[-0.010, 0.063]	92.3%
Total media (t – 1)	0.025	[0.003, 0.047]	98.5%
Between-subject effects			
Intercept	-1.279	[-1.601, -0.937]	>99.9%
Strength	0.103	[0.027, 0.185]	99.7%
Esteem	-0.073	[-0.151, 0.001]	97.5%
In-party talk	0.307	[0.207, 0.411]	>99.9%
Total talk	-0.250	[-0.354, -0.160]	>99.9%
Total media	0.371	[0.305, 0.436]	>99.9%
Republican	-0.051	[-0.144, 0.052]	83.8%
Gender (woman)	-0.061	[-0.139, 0.021]	92.3%
Issue alignment	0.070	[0.036, 0.108]	>99.9%
Race/ethnicity (White)	-0.030	[-0.144, 0.095]	70.9%
Race/ethnicity (Hispanic)	0.096	[-0.056, 0.246]	89.0%
Race/ethnicity (Black)	-0.014	[-0.157, 0.141]	61.5%
Age	0.009	[0.000, 0.018]	98.4%
Voted before	-0.105	[-0.176, -0.039]	>99.9%

Note: "Post. Prob." refers to posterior probability, which readers can interpret approximately as 1 – p.

Chapter 10. Discussion

Why are political allegiances so resistant to change? What are the consequences of a communication environment with so much choice and access to partisan information? This study has engaged with some of the enduring questions of political communication research. In this context, its contribution is necessarily modest. I have argued that, as a theoretical matter, communication scientists should consider stability, especially stability of attitudes, as a potential outcome of communication. In other words, stability is not necessarily the default state people revert to when there are no effects, but rather stability is perhaps one of the effects which we should study. I am far from the first to make this suggestion, but it is one that bears repeating. To advance this goal, I have discussed in both theoretical and empirical terms what it means to study stability in addition to change. The typical media effect that occupies the focus of most communication research is what I have referred to as development (a la Nesselroade, 1991). Development refers to changes in the mean level of some variable — generally either the amount of communication or a psychological construct — that endures at least long enough to observe and predict. When we want to study stability, we need to focus on the types of variation besides development.

In the longitudinal setting, I have broken this into time-dependent and net variability. Time-dependent variability refers to how much today's deviation from the norm is determined by yesterday's deviation from the norm. If someone spent much more time than normal talking about politics yesterday, would we expect them to talk politics a different amount than normal today? If yes, this is an example of time-dependent variation. If not, this suggests an ability to

self-regulate the behavior, which is a form of stability. Net variability, on the other hand, refers to the frequency and magnitude of those deviations from the norm. A person who typically avoids all discussion of politics but occasionally engages in many hours of it is an example of someone with high net variability and low time-dependent variability (because they immediately return to their norm of no discussion). By contrast, someone who reads the morning newspaper for 15 minutes on the bus every day but always avoids the news the rest of the day has low net variability and low time-dependent variability. Sometimes the bus ride may be quicker or slower than normal, but never by much (net variability). And their typical amount of consumption does not change permanently because the typical bus ride is predictable and a long or short ride does not precipitate a subsequent increase or decrease in news consumption outside of the morning commute.

Another challenge to research that I have discussed at length is the observation that key constructs for political communication research — like media use, political discussion, and partisanship — appear by some analyses to be extremely stable over time. Normative concerns aside, if we stipulate that it is true that all these constructs rarely vary, then researchers have a practical problem. Most scientific theories relevant to quantitative communication researchers concern changes in the constructs under analysis. Even if the theories are accurate descriptions of reality (what would happen to the dependent variable if the independent variable(s) changed), they would be nearly impossible to investigate and their relevance limited if such changes rarely occur in the real world. The situation is not so dire, however. I show that although a degree of stability is indeed common to all these constructs, the evidence for extreme stability is limited and often methodologically problematic. The psychometric methods used to calculate high

stability rely on assumptions that are likely violated in ways that overstate stability for these constructs. Furthermore, such analyses dismiss both time-dependent and net variability as uninteresting; analytically, these methods are oriented around removing this type of variation altogether to focusing only on development. Although I will not dive into the statistical details here, I have laid out a set of best-practices to take this understanding of types of variability and carry it over into quantitative research design and analysis.

In the case of partisanship, I argue that the theoretical focus has been on the wrong target: the party one prefers rather than the degree of affiliation with that party. Using the social identity approach as background, I describe how party preference is just one dimension of partisanship. In addition to the strength of affiliation, I introduced collective self-esteem as a relevant dimension of identity that is rarely discussed in the context of politics. Collective self-esteem should be causally related to the strength of one's identity insofar as low collective self-esteem should reduce identity strength. The data collected for this dissertation suggest this is indeed the case. As for stability, theory suggests that party preference is most stable, followed by strength of partisanship, and collective self-esteem most variable. The data likewise suggest that collective self-esteem exhibits less between-person variance (people differ from each other less) but more within-person variance (they change more from one day to the next) than identity strength. This is consistent with the notion that collective self-esteem is a sensible target for research purposes; it is both an important component of identity and one that is relatively malleable. Indeed, the apparent mechanism of much experimental social identity research outside the political domain involves manipulating collective self-esteem by making groups seem to have higher- or lower-status.

The field of political communication investigates two basic types of relationships between communication and partisanship. One, which I refer to as the media effects approach, investigates whether and to what extent communication changes partisanship or other political attitudes. The other, best represented by selective exposure theory, instead treats communication as the outcome variable that is determined by partisanship or political attitudes. The approaches do not necessarily see one another as mutually exclusive and indeed one implication of the two approaches is that the two phenomena may feed off one another. One theory that is novel in part for linking the two is the Reinforcing Spirals Model (RSM; Slater, 2007, 2015). Although part of the RSM's heuristic appeal is how it suggests the possibility of constantly-increasing extremity in both attitudes and communication behavior, it also provides a framework for explaining why such a process may be the exception rather than the norm. The RSM postulates that due in part to external forces and competing interests, people are not motivated to engage in any more communication than is necessary to maintain their identities and attitudes. The reason one would need to engage in communication for this maintenance, per the RSM, is that otherwise identity *threats* — negative information, hostile interactions with out-group members, and so on — would erode the strength of the identity. I have suggested an elaboration on this, such that identity threats are not the only reason people need to communicate to maintain their identities. Instead, identities intrinsically require certain regular practices, like communication, to remain strong. If never prompted to actively think about an identity, how could it be central to one's self-image, as is the definition of a strong identity? The categorical aspect, such as whether one is a Republican or Democrat, may persist in the absence of communication or other identity-

affirming activities, but these activities seem essential to maintaining the importance of the identity in one's life.

Empirical Study Results

For the empirical study, I offered several hypotheses and research questions. H1 addressed the effect of in-party communication on the variability of identity strength, predicting it is part of the process of self-regulation. The RSM argues that people are driven to maintain homeostasis. That is, there is no inherent motivation to make one's own identity more (or less) extreme, but people do generally aim to resist change. A person encounters various potential challenges to his or her beliefs and identities in daily life, both in terms of other issues competing for attention and the social environment providing potential threats to those beliefs and identities. Communication should serve to alleviate those threats and to satisfy basic psychological needs to engage with issues of importance. Maintaining homeostasis can take two forms in my framework. One, restricting net variability, concerns the ability to quite literally maintain the same level of identification over time and not waiver on occasion. The other, reducing time-dependent variability, is about the ability to take those momentary shifts in identification and return to normal relatively quickly.

A consistent pattern in the results was no effects on time-dependent variability for any dependent variable. Furthermore, modeling results indicate identity strength (and the in-party communication variables) exhibited no time-dependent variability, on average. In other words, whether a participant departed from his or her usual level of identity strength could not generally be predicted from whether the participant departed from his or her usual level the day before.

Since the hypothesis was that in-party communication would push time-dependent variation towards zero, it is difficult to say whether there could be an effect when it appears that zero is the usual amount among these participants. It is worth noting that if one was to model identity strength with no other predictors, then there would be substantial positive autocorrelation; but this carry-over of changes from one time to the next is completely explained by the other variables rather than intrinsic to strength itself. This means it is probable that communication has a role in this hypothesized process, but the data do not support a conclusion about precisely what kind of communication. As an interpretational note, I find the fit of time-dependent variability into communication theory less clear than its counterpart, net variability. This terminology originates in psychology, where the interest in time-dependent variability has been focused on the regulation of one's emotions. In that context, it is clear that both positive and negative time-dependent variability is dysfunctional; if high, then a person lacks the ability to recover emotionally from environmental stimuli. If low, a person rapidly shifts between too-positive and too-negative moods in a toxic cycle of over-compensation. For the variables under study in this dissertation, it is less clear what the substantive implications are for the presence of positive or negative autocorrelation. It is worth noting that collective self-esteem, the sole variable with non-zero autocorrelation, is the variable most similar to an emotional state. The existence of autocorrelation is also the most substantively interpretable: A person whose collective self-esteem is harmed takes some time to get back to normal.

Another prediction, part of H1, was that in-party communication would suppress net variability of identity strength. This was observed for in-party media and in-party discussion. The net effect of in-party discussion, however, depends somewhat on the amount of discussion in

general; discussion appears to greatly increase net variability of identity strength, but in-party discussion in effect helps to cancel it out.

In addition to variability, H2 takes on the media effects approach, predicting increases in identity strength (i.e., development) as a result of in-party communication. This was observed for both in-party discussion and in-party media. The same prediction was offered with collective self-esteem as dependent variable and the same pattern was observed. When it comes to the determinants of in-party communication, the data provide fewer answers. H6 predicted that increases in identity strength would increase in-party communication. This relationship was not observed for discussion or media, although for discussion there was a statistically weak effect in the positive direction. Although I will not interpret these results as evidence in favor of the null hypothesis, they are consistent with the notion that media effects predominate over selective exposure, at least on the within-person level. Communication research has long considered both the extent to which media and interpersonal discussion might change attitudes as well as how attitudes affect the quantity and qualities of communication. Generally speaking, research has long shown that there is typically an association between one's attitudes and the content of their communications such that people seem to be exposed to more congenial views than expected by chance (e.g., Sears & Freedman, 1967). Of course, whether this occurs because the communication changed one's attitudes or because one's attitudes caused the choice of communication cannot be determined by simply noting this correlation. In the laboratory, there is little doubt that people show a preference for congenial information (e.g., Hart et al., 2009; Taber & Lodge, 2006). Yet the laboratory can show effects of communication on attitudes as well (e.g., Hovland, Lumsdaine, & Sheffield, 1949). Knowing that both selection and influence *can* occur is

important, but insufficient to explain how media diets appear congenial and social networks tend to be homophilous. These results comport with the recent finding — also among college students — that the partisanship of political discussants appears not to be driven by partisan identity but other social motivations (Minozzi, Song, Lazer, Neblo, & Ognyanova, 2020). They are also consistent with the claim that media use is highly stable, which implies that it is a behavior not particularly responsive to changes in other variables (Scharkow, 2019). Nonetheless, the challenge of measuring communication and the dangers of linear models with such a highly skewed dependent variable could plausibly explain the null findings as well. This study does not and could not resolve this decades-old, field-spanning question.

In-party media use did promote subsequent in-party discussion, even though the reverse was not indicated by the data. This suggests a potential mediating path through which in-party media use may additionally affect those psychological variables. Given the lack of prior hypothesis and the many assumptions required, I did not perform any formal mediation tests to probe this further (but if both paths have statistically significant coefficients, this offers a conservative test of mediation; Hayes & Scharkow, 2013). Although the data cannot speak to opinion leadership per se, there is a parallel with these findings to the claims of *Personal Influence* (Katz & Lazarsfeld, 1955). Katz and Lazarsfeld argued that even when media use does not directly produce attitude change — though it does so in my data — it can do so indirectly by prompting interpersonal discussions, which are the proximal cause of attitude change. That being said, the comparability is limited; I have measured the amount of discussion, not the content of it, and my data do not shed any light on the extent to which these media to discussion effects may be concentrated in opinion leaders. But more recent work has continued to investigate this

same pathway from media to discussion (e.g., Shah, Cho, Eveland, & Kwak, 2005). Discussion has also been suggested as the second "O" in the O-S-O-R model of communication (McLeod, Kosicki, & McLeod, 1994) in which media serves as stimulus ("S") and discussion can be part of the subsequent orientation that precedes a response ("R"). The communication mediation model (Sotirovic & McLeod, 2001), itself partly inspired by the O-S-O-R model, and its derivatives and developments (Shah et al., 2017) also explicitly place discussion as a mediator between media and (often behavioral) outcomes. This study was not designed to engage with this literature, but the finding fits into a long tradition of communication models and so should not be cast off as merely an anomalous result.

Furthermore, I predicted that collective self-esteem would boost identity strength, and this was supported by the analyses. This study was not meant to be a rigorous test of the nuances of social identity theory, but the results do support this tenet of the social identity approach. For the purposes of the underlying communication theory, collective self-esteem provides another potential explanation for why someone might be motivated to engage in communication for identity maintenance. Likewise, it is suggestive of a causal pathway for how communication can influence identity strength: making the group seem like a better one to belong to. That there are direct effects from communication to identity strength, however, suggests there may be another pathway as well. Another potential explanation for this is that some of the feedback between collective self-esteem and identity strength occurs too quickly to be captured even with once-daily measurements. RQ2 asked whether collective self-esteem influences communication, but the results did not show this relationship. This argues against the notion of communication being used intentionally to alleviate threats to identity. The intentionality aspect is important; the fact

that communication appears to support collective self-esteem means that in practice, the data provide evidence for this hypothesized protective effect that is part of the RSM. But this effect may be incidental and not part of a conscious process in which a person changes his or her communication routine to respond to identity threat. Such disconnects between the reasons for communication and the actual outputs are not unheard of in prior research (Z. Wang & Tchernev, 2012).

RQ4 asked whether communication that does not come from in-party sources has any effect on variability of identity strength. Indeed, as mentioned, political discussion seems to promote much greater variability and in-party discussion can only partially cancel out this effect. This is in spite of the fact that it had no effect on development of identity strength; we can take from this that such "dangerous" discussion can throw people's sense of partisan identity off track, but not predictably one way or another. Conversely, media use appears to affect development of identity strength — reducing it — but had no apparent effect on variability, suggesting its effects are more predictable than discussion. One way to think about these effects on variability is in terms of change that cannot be explained. The results for discussion clearly suggest that discussion changes identity strength in that this is the literal meaning of variability. But the change is not in the mean, so the change is in some sense not systematic. Does this mean the change is random? Probably not. In my view, it implies there are factors not adequately captured in the model that determine the outcome of such discussions. These may be measurable variables, like the content of the discussion. Importantly, one can only guess at the extent to which the discussions participants reported were characterized by (dis)agreement. Research on political disagreement has produced conflicting findings in part due to inconsistent measurement

and conceptualization of disagreement (Klofstad, Sokhey, & McClurg, 2013). My measurement approach is consistent with how some have purported to have measured disagreement — assuming that the partisan attachment of the other person is a surrogate for agreement — but I make no such claim here. Discussion with apparently similar people, at least when they are part of a close relationship, often includes disagreement (Morey, Eveland, & Hutchens, 2012). The potential roles of disagreement and its interaction with partisanship, identity strength, and other individual factors are too numerous to catalog here, but this may be an area where some of this unexplained variation can become explained.

Limitations

There are several limitations to consider in evaluating the empirical study. First is the sample: Students were chosen not for their theoretical relevance, but rather due to their availability at an affordable cost. Findings demonstrated in student samples often generalize to other populations, but not always (see, e.g., Krupnikov & Levine, 2014). Even a highly unrepresentative, but more diverse sample would likely have been preferred in terms of generalizability (Clifford, Jewell, & Waggoner, 2015; Coppock, 2018; Coppock & McClellan, 2019). It is plausible that students are *more* amenable to testing theoretical questions that concern the malleability of stable psychological constructs due to their still-in-progress socialization, but even if true this would still be an example of their non-generalizability. None of this is to say that nothing can be learned from a student sample or that it should not have been used given the costs and benefits, but it is nonetheless a weakness of the study that should be borne in mind when evaluating its results. Although the somewhat novel measures used in this study make it difficult to do apples-to-apples comparisons with prior work, it may be informative to compare the

strength of partisanship data to the British data used by Huddy, Bankert, and Davies (2018; see also Bankert et al., 2017). Using the same scaling procedure they used, the distribution of partisan strength appears quite similar in these data to their British general population sample. Although a coarse comparison, they indicated that using conventional measures of partisanship, the British exhibit a similar level of partisanship in aggregate to the United States, suggesting these students may not be particularly more or less partisan than non-students. There is not stability data readily available from that British sample — difficulties quantifying stability at the construct level notwithstanding — but I am able to reconstruct the path model used by Huddy and colleagues (2018) that they used to make claims of stability in identity strength. The 1-lag autocorrelation in this study is of quite similar magnitude, albeit slightly greater, than they observed. This is to be expected given the shorter time scale between observations; I defined 1 "lag" in this informal analysis as a week in my data to maximize the spacing between data to best approximate Huddy et al.'s procedure. The difference in timescale (the British data has several weeks pass in between its irregularly-spaced observations) makes strict comparability difficult but this is at least suggestive that this student sample does not exhibit much higher instability compared to a more diverse population in another English-speaking democracy.

As discussed extensively in an earlier chapter, the measurement of communication — especially media use — is one of the major challenges in the discipline. I will not use the space here to review all the difficulties involved in generating accurate self-reports of media use. Less evidence exists regarding the accuracy of self-reported measures of discussion, but this may be driven in part by a lack of readily available ground-truth data like there is for some forms of media use. The convergent validity evidence I presented, however, suggests my measurement of

discussion may have less error than my measure of media. Participants' daily reports of discussion correspond rather closely with their initial estimates of how many days per week they talk about politics as did the overall distribution of in-party and out-party discussion. The daily measures of media were not nearly as closely correlated with a separate measure of media use in the initial survey. This is worth bearing in mind when considering that in-party discussion, but not in-party media, was found to increase identity strength and collective self-esteem. A construct measured with more (random) error is more likely to be a false negative in a multivariate analysis (Achen, 1983). Another important point about these difficulties in measuring communication is that although there is little doubt that these measures contain significant error, and are probably biased upward (Prior, 2009), no research to date shows that these problems bias hypothesis tests in any way besides attenuating estimates.

The validity of the findings depends crucially on the appropriateness of the statistical models used. This includes both the specification and estimation techniques used as well as the methods used for missing data imputation. Some robustness testing is included in the Appendix. As far as missing data is concerned, findings are little changed when using another prominent software package for R (called "mice"), which uses a different underlying method than the one used in this dissertation. Decisions about inclusion and exclusion of participants are difficult and guidance in the literature is not clear. Results using different decision-making processes are presented in Appendix A2 and show largely similar patterns regardless of inclusion/exclusion criteria, although more exclusions generally attenuate estimates. I further probe some of these uncertainties in Appendix A3 using Monte Carlo simulation. Under mild violations of the assumptions regarding the causes of missingness — complete case analysis is generally only

assured to be valid when respondents fail to respond for reasons that are completely random — the approach I used proves most trustworthy. When those assumptions are not violated, my approach can produce false positive results, but the error rate could be controlled by simply imposing a higher standard for statistical significance. Most statistical results that have bearing on the conclusions would withstand such adjustments. Approaches that rely solely on the observed data or exclude more of the missing observations are more susceptible to bias from assumption violations and under some conditions have extremely poor statistical power. The most problematic type of violated assumption is when those who do not respond differently to the "treatments" (i.e., the predictors) than those who do, in which case the results can be summarized as: more imputation is better. Even in the absence of violated assumptions, complete case analysis will produce more variable estimates. This means not only does one expect wider confidence intervals from complete case analysis, but it is statistically likely to see some apparently discrepant results between complete cases analysis and methods based on imputation even if there would be no difference between them in the long run (i.e., in a world where the study is conducted many times and the results are averaged). It is unfortunately not clear without the benefit of running many studies whether results are discrepant due to chance or because one or the other model is reflecting a true bias.

Setting aside missing data, the models used themselves are not beyond reproach and rely on certain assumptions for interpretation. For effects on development, there are somewhat reasonable assumptions required to interpret estimates causally; confounding is only possible from unmeasured variables that change over time. For between-subject comparisons like those used to test hypotheses regarding variability, there is a greater threat of confounding more

analogous (but not equivalent) to cross-sectional analysis. The best way to get around these issues is to use experimental manipulations, but these provide no free lunch either; forced exposure to media, for instance, does not produce the same effects as voluntary exposure that occurs in the real world (Stroud, Feldman, Wojcieszak, & Bimber, 2019). These problems are not insurmountable, however, and methodological triangulation can be a way to address the limitations of any single approach. Another issue that is likely to apply to analysis of communication as dependent variable no matter the design or measurement used is that the distribution of communication behavior is not particularly amenable to analysis via linear models. This is not a serious problem for estimating changes in terms of mean levels (i.e., development) — there are other kinds of models, like those intended for count data, that are available. But for estimating effects on variability, particularly net variability, use of non-normal models is not well-studied and there are not typically variance parameters to analyze as there are when assuming normally distributed errors. It is not uncommon for models with assumption violations to still produce substantively correct results, but it is nonetheless a source of uncertainty for models presented here that have communication variables as the outcome. For this reason, I have not put as much emphasis on those results.

In the context of the RSM, Slater (2007) says, "the fact that one has longitudinal data is no guarantee that one can detect spiral processes if the measurement lags do not reflect a good understanding of the underlying processes" (p. 286). A strength of this study is its high resolution; that is, the frequent measurements of the key variables. But implicit in the design and analysis is that daily measurements reflect the timescale on which the hypothesized effects occur. Logically speaking, to the extent that communication has an effect on some psychological

variable, it should occur almost instantaneously. On the other hand, while such an effect may in truth occur right away, whether the effect (a) is large enough to detect and (b) persists long enough to detect is not so obvious. Effects may therefore need to accumulate to be noticeable and complex dependencies are possible, such as a requirement of several consecutive days of communication at some level before (detectable) effects occur. Standard statistical tools are not well-suited to deal with these possibilities. The existence of some effects in the models may be seen as evidence for the models' appropriateness, but this is of course tautological and it does not resolve whether null effects are interpretable as an absence of effect or something more complicated. Another issue to consider is that to assess the long-term durability of effects and the extent to which effects may accumulate over larger timescales, the daily structure of this study may not be appropriate. One way to combine the benefits of the intensive longitudinal design with the desire to assess longer-term processes is a so-called "measurement burst" design (Nesselroade, 1991; Ram & Gerstorf, 2009). In such a study design, participants are subject to "bursts" of measurements on daily or even more frequent time scales on a periodic basis, like months or years. For instance, researchers could do a week of daily measurements every 3 months in order to have data that reflects the multiple timescales of interest.

On the conceptual level, I have treated the communication variables as if they have a clear and stable meaning. For instance, discussion with supporters of the same party should produce greater collective self-esteem and identity strength because of a mixture of simple priming (thinking about one's identity) and because it should foster a sense of affiliation due to the content and context of the conversations. This design cannot parse which of those two explanations, if either, would explain the observed effects. And it is of course true that not all

same-party discussions will consist of agreement over shared values and attitudes. Conversely, discussion with supporters of the other major party may not always be objectively threatening or contain explicit disagreement. People likely make efforts to limit their discussions with out-party supporters to subjects on which disagreement is less likely or less profound (Cowan & Baldassarri, 2018). It is also plausible that even equivalent discussion content may have qualitatively different effects based solely on the identity clash of the participants due to different interpretations of what is said, the difficulty of confronting hard truths in the context of intergroup competition, or just plain anxiety about what disagreements might occur as the conversation moves forward. This is in part the reason why the measures used in this study focus on the source of communication rather than the content. This is all to say that the effects and non-effects observed here refer to these communications as they happened, but what exactly happened — what was discussed, with what level of emotion, with how many people — is not clear. Whether the character of these communications would be similar for other sample populations is another area of uncertainty.

This study has had little to say about political independents. That group has puzzled and frustrated many researchers of American politics for both theoretical and empirical reasons, seeing as many of them both self-describe as not affiliated with a party despite acting a lot like affiliates of a party (e.g., J. Dennis, 1988; Keith et al., 1992). For this study, I classified as partisans those who initially claimed to be either independent or a supporter of a major party but later said they are closer to a major party, given prior work on this question (Petrocik, 2009). The independents who "lean" to one party or another are qualitatively different to those who do not, particularly in terms of their greater engagement with politics (e.g., media use, voter turnout;

Klar, 2014; Petrocik, 1974). In general, it is best to consider independents as a more heterogeneous group than partisans. The present study only included leaners, given the prior work on how they are often misclassified as non-partisans. The clearest difference among them in these data is in their level of identity strength, which was approximately a standard deviation lower than those who did not initially select another party. The difference in strength of identification is larger than the difference in their ideologies as measured by issue positions. They use political media and talk about politics just as much as non-leaners, but they do spend slightly less of their time talking to same-party supporters. Whether this is cause, effect, or both, of their lower level of identification is of course not clear.

Previous research has suggested that political independents, leaners and non-leaners alike, can conceptualize their independence as an identity (Klar, 2014). Whether this is a *social* identity is less clear; at least in the United States, to be independent-minded is generally considered a positive personality trait, making independence an appealing target as a *personal* identity. Petrocik (2009) argues that leaners should be understood "as partisans who are engaging in a self-presentation" (p. 572) and that true independents have historically made up about 10–15% of the American public. Leaners generally look like weak partisans in terms of their ideology and party loyalty, although they tend to be more engaged in politics. Klar's (2014) findings suggest that a likely difference between leaners and those who do not choose to label as independent and instead say they are weak partisans is that independence is a relatively important part of the self-concept of a leaner. Klar and Krupnikov (2016) also show that independents of all stripes share some disdain for partisan politics regardless of party, consistent with Clifford's (2016) finding

that the strength of party affiliation is determined partly by a person's general tendency to value group membership.

In my data, when it comes to evaluation of the parties, the leaners distinguish themselves from those who initially chose a party in terms of their in-party evaluation. In a feeling thermometer measure taken in the initial survey, non-leaning partisans gave their own party a rating of 75 (out of 100) on average, compared to 55 for lean partisans. Compare this to the gap in evaluations of the other party, which non-leaning partisans rated an average of 27 while lean partisans rated it an average of 38. The difference in evaluations is nearly twice as great for the in-party evaluation compared to the out-party. This is consistent with the results reported using data from the American National Election Study by Klar, Krupnikov, and Ryan (2018) using a slightly different operationalization of partisanship. One way to conceive of this distinction is negative partisanship, which is when political behavior is driven more by dislike of one party rather than liking towards another (e.g., Medeiros & Noël, 2014). It is certainly true that leaning and non-leaning partisans are in more agreement in terms of their evaluation of the opposing party, but it should also be considered that the typical leaning partisan does not feel particularly negatively towards that party either. True independents, on the other hand, tend not to have consistent party preferences and so may have yet other ways of sense-making in the political realm. New research provides ways of trying to measure negativity towards individual parties independently of positivity, suggesting a way to incorporate the views of independents in future designs like this one (Bankert, 2020).

Future Directions

This discussion of limitations lends itself to consideration of the best way to move forward. There are several clear pathways to improve upon the current study. First, a more diverse sample at minimum and ideally a sample representative of a relevant population (e.g., U.S. adults) would be used to establish whether any aspect of the results of this study were peculiar to students taking communication courses at Ohio State University. Another major area for improvement is the measurement of communication, especially of media use. In the best conditions, measuring media use is difficult; in an intensive longitudinal design, it is made more difficult by the need to reduce the burden on respondents. Although imperfect, methods for passive tracking of media use could be a very valuable tool for these designs since they avoid reporting biases while also reducing the daily burdens of response. Passive tracking of political discussion is far more complicated, but one does not have to focus on this and I have shown there is reason to believe self-reports of discussion may be more accurate than for media.

To add to the robustness of causal claims, experimentation is possible within intensive longitudinal designs. For instance, participants could be sent daily newsletters or pinged with periodic news alerts to prompt exposure to different kinds of media. Such manipulations could be both between- and within-subjects. It is likely there would need to be a control condition that involves some non-political news. Manipulating identity strength or collective self-esteem by inducing identity threat, on the other hand, could be more difficult, although collective self-esteem would likely be easier. For instance, in the context of racial identity, Craig and Richeson (2014) as well as Bai and Federico (2019) confronted White participants with information about

Whites potentially losing their majority status in the United States. This could trivially be modified to focus on partisan advantage instead. It should also be mentioned that in some cases, the content of communication and objective information about group status may be one and the same, although trust and other responses may depend on the source. For the most obvious example, election results are typically learned through media and may have major implications for group identity, but it is likely that the information is more important than the source in that situation. Future work should also consider using the content of news — widely available thanks to several digital initiatives — as a way to parse out the effects of content versus source.

There may be effects involved in these processes that are not adequately accounted for by the analyses presented. For instance, implicit in the within-person analytical approach is the assumption that changes in a variable have the same effect no matter what the previous value of the variable was. In concrete terms, this means that the analyses assume that a 5 minute increase in media use has the same effect regardless of whether yesterday's media use was 0 minutes or 120 minutes. This is somewhat addressed in these particular analyses because of the log transformation applied to the communication measures, but only indirectly. Future work can explore in a more targeted way whether these within-person effects differ depending on the absolute level of the variables. To continue the example, one could include an interaction between the within-person deviation and the person-specific average to represent the extent to which those changes might depend on the typical levels of a variable. A plausible hypothesis in the case of communication is that changes in communication might be more impactful among those who do relatively little of it. Another, related idea may be that increases and decreases in the variables have different effects; that is, perhaps a decrease in self-esteem triggers a different

response than an increase. There are relatively new analytic models (Allison, 2019) and software (Long, 2020) available to test these types of questions as well.

Another avenue for subsequent research is to explore other socially important identities, such as race, although there will be special considerations and important theoretical work necessary to adapt this design to those contexts. A full explication is outside the scope of this dissertation, but in particular it would be important to establish what kind of communication is relevant to racial and ethnic identity. The shorthand used for this study — talking with members of the same group or engaging with media that aligns itself with the group — is probably not always appropriate in the racial and ethnic identity context. One useful example of an approach to media and ethnic identity is work by Saleem and colleagues (2019), who found that American Muslims exposed to negative news coverage of Muslims had less strength in their identification as Americans. That study is also a good example of how when studying racial and ethnic identity, the most straightforward approach may be to focus on a single group at a time to better capture the types of relevant communications and issues.

Even if one sought a relatively straightforward replication of this study, there are some lessons learned worth consideration. One of these would be to try to verify the quality of the contact information provided by participants; upon inspection, several participants seem to have provided invalid email addresses and/or phone numbers. In some cases, these could potentially be checked immediately before forms are submitted, like if a phone number has the incorrect number of digits or an email address lacks an @ symbol. In others, an investigator can check for error messages related to the attempted sending of emails and text alerts in order to try to reach

out and correct the issue. Researchers may also wish to host formr (the survey software) on their own servers; on a few occasions, the developers' server in Germany had unexplained outages for several hours that prompted respondents to ask for help and potentially not provide data. *Caveat emptor*, however, since such failures are probably more likely on non-professionally-managed servers.

One issue in the interpretation of non-response was that participants were recruited in a way that informed them that they had no obligation to continue beyond the initial survey; they were told that they could earn a designated number of credits for completing that survey and could optionally continue participating for additional credits. Making this opt-in/opt-out explicit in the first survey would make it easier to understand why some participants never again responded, although it could potentially screen out commitment-wary participants as well. Relatedly, those seeking to address non-response would do well to consider whether additional measures would be valuable, like another reminder at mid-day or towards the evening for those who did not complete their daily surveys in the morning. Some kind of additional intervention for those who have not responded for several days in a row might also be worthwhile.

There are a few issues related to the measurements of identity strength and collective self-esteem worthy of consideration as well. One is the choice of the number of items shown/omitted per daily survey. My decision to randomly select 4 for each scale was fairly arbitrary, so it would be worthwhile have a more empirically-based decision in the future. Even one of the premises for showing a subset of the scales is relatively untested: Little to no research gives guidance on the likelihood that respondents would respond differently due to remembering their past

responses, how long they are likely to remember, and what the consequences are of this kind of invalidity. If the likelihood of remembering is greater or more consequential than previously thought, then the impact of scale length on reliability becomes more important since one would want to show an even smaller subset of the items to reduce the memory issues. On a more pragmatic note, the item response theory analysis indicates that two of the items on the collective self-esteem scale do not add much information. These are the final two in the list in Appendix A1, "other people think that [Republicans/Democrats] are unworthy" and "[Republicans/Democrats] are not respected by the broader society," both reverse-coded items. This should be considered in light of recent psychometric research questioning the use of reverse-coded question wording (Suárez-Álvarez, Pedrosa, & Lozano, 2018), although it should also be noted that two other such items were not so problematic in these data.

Conclusion

This study took aim at broad and enduring questions faced by those who study political communication. Doing so comes at a cost — one study can only do so much to move the ball on these issues that have received so much attention. The payoff, of course, is getting into the dialogue on these important issues. Perhaps the most important contributions of this project are theoretical. The field has often noted that what constitutes an "effect" is more than just qualitative shifts in a dependent variable. Recognition that lack of change is itself an outcome of the communication runs at least as far back in time as Columbia's Bureau of Applied Social Research (1948). But for quantitative research to use this long-known theoretical insight, statistical tools and appropriate study designs are necessary. Along those lines, then, another

contribution is the description of those tools. And the empirical results add more evidence to the literature on the relationship between political identity and political communication in everyday life. The evidence here suggests that at least on a day-to-day basis, communication affects identity in the traditional media effects sense while those shifts in identity strength are not particularly influential on day-to-day communications, consistent with a model of communication as highly routinized and subject to exogenous forces. Communication can also both entrench and destabilize identity, depending on the kind of communication. Peering more closely at partisan identity, the evidence further suggests the value of considering the components of that identity separately, particularly the strength of one's identity and collective self-esteem.

Bibliography

Abramowitz, A. I. (1988). An improved model for predicting presidential election outcomes. *PS: Political Science & Politics*, *21*, 843–847. https://doi.org/10.2307/420023

Abrams, J. R., Barker, V., & Giles, H. (2009). An examination of the validity of the Subjective Vitality Questionnaire. *Journal of Multilingual and Multicultural Development*, *30*, 59–72. https://doi.org/10.1080/01434630802369452

Abrams, J. R., Eveland, W. P., & Giles, H. (2003). The effects of television on group vitality: Can television empower nondominant groups? *Annals of the International Communication Association*, *27*, 193–219. https://doi.org/10.1080/23808985.2003.11679026

Abrams, J. R., & Giles, H. (2007). Ethnic identity gratifications selection and avoidance by African Americans: A group vitality and social identity gratifications perspective. *Media Psychology*, *9*, 115–134. https://doi.org/10.1080/15213260709336805

Achen, C. H. (1975). Mass political attitudes and the survey response. *The American Political Science Review*, *69*, 1218–1231. https://doi.org/10.2307/1955282

Achen, C. H. (1983). Toward theories of data: The state of political methodology. In A. W. Finifter (Ed.), *Political science: The state of the discipline* (pp. 69–94). Retrieved from http://archive.org/details/politicalscience00hodo

Achen, C. H. (2000). *Why lagged dependent variables can suppress explanatory power of other independent variables*. Presented at the Annual Meeting of the American Political Science Association, Los Angeles, CA.

Alemzadeh, S., Niemann, U., Ittermann, T., Völzke, H., Schneider, D., Spiliopoulou, M., … Preim, B. (2020). Visual analysis of missing values in longitudinal cohort study data. *Computer Graphics Forum*, *39*, 63–75. https://doi.org/10.1111/cgf.13662

Allison, P. D. (2009). *Fixed effects regression models*. https://doi.org/10.4135/9781412993869.d33

Allison, P. D. (2019). Asymmetric fixed-effects models for panel data. *Socius*, *5*, 1–12. https://doi.org/10.1177/2378023119826441

Althaus, S. L., Bramlett, B. H., & Gimpel, J. G. (2012). When war hits home: The geography of military losses and support for war in time and space. *Journal of Conflict Resolution*, *56*, 382–412. https://doi.org/10.1177/0022002711422340

Alwin, D. F. (2007). *Margins of error: A study of reliability in survey measurement*. Hoboken, N.J: Wiley-Interscience.

Arslan, R. C., Walther, M. P., & Tata, C. S. (2019). formr: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R. *Behavior Research Methods*. https://doi.org/10.3758/s13428-019-01236-y

Bai, H., & Federico, C. M. (2019). Collective existential threat mediates White population decline's effect on defensive reactions. *Group Processes & Intergroup Relations*, 136843021983976. https://doi.org/10.1177/1368430219839763

Baird, B. M., Le, K., & Lucas, R. E. (2006). On the nature of intraindividual personality variability: Reliability, validity, and associations with well-being. *Journal of Personality and Social Psychology*, *90*, 512–527. https://doi.org/10.1037/0022-3514.90.3.512

Bakshy, E., Messing, S., & Adamic, L. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*, 1130–1132.

Bankert, A. (2020). Negative and positive partisanship in the 2016 U.S. presidential elections. *Political Behavior*. https://doi.org/10.1007/s11109-020-09599-1

Bankert, A., Huddy, L., & Rosema, M. (2017). Measuring partisanship as a social identity in multi-party systems. *Political Behavior*, *39*, 103–132. https://doi.org/10.1007/s11109-016-9349-5

Barber, M., & Pope, J. C. (2019). Does party trump ideology? Disentangling party and ideology in america. *American Political Science Review*, *113*, 38–54. https://doi.org/10.1017/S0003055418000795

Bartels, B. L., Box-Steffensmeier, J. M., Smidt, C. D., & Smith, R. M. (2011). The dynamic properties of individual-level party identification in the United States. *Electoral Studies*, *30*, 210–222. https://doi.org/10.1016/j.electstud.2010.11.002

Bartels, L. M. (2000). Partisanship and voting behavior, 1952-1996. *American Journal of Political Science*, *44*, 35–50. https://doi.org/10.2307/2669291

Bartels, L. M. (2014). Remembering to forget: A note on the duration of campaign advertising effects. *Political Communication*, *31*, 532–544. https://doi.org/10.1080/10584609.2014.956198

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software*, *67*. https://doi.org/10.18637/jss.v067.i01

Bauer, R. A. (1959). Comments on “the state of communication research.” *Public Opinion Quarterly*, *23*, 14–17. https://doi.org/10.1093/poq/23.1.14

Bell, A., & Jones, K. (2015). Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods*, *3*, 133–153. https://doi.org/10.1017/psrm.2014.7

Bello, J., & Rolfe, M. (2014). Is influence mightier than selection? Forging agreement in political discussion networks during a campaign. *Social Networks*, *36*, 134–146. https://doi.org/10.1016/j.socnet.2013.06.001

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 6, pp. 1–62). https://doi.org/10.1016/S0065-2601(08)60024-6

Bennett, W. L., & Iyengar, S. (2008). A new era of minimal effects? The changing foundations of political communication. *Journal of Communication*, *58*, 707–731. https://doi.org/10.1111/j.1460-2466.2008.00410.x

Berelson, B. (1959). The state of communication research. *Public Opinion Quarterly*, *23*, 1–2. https://doi.org/10.1086/266840

Bergami, M., & Bagozzi, R. P. (2000). Self-categorization, affective commitment and group self-esteem as distinct aspects of social identity in the organization. *British Journal of Social Psychology*, *39*, 555–577. https://doi.org/10.1348/014466600164633

Bland, J. M., & Altman, D. G. (1996). Statistics Notes: Measurement error and correlation coefficients. *BMJ*, *313*, 41–42. https://doi.org/10.1136/bmj.313.7048.41

Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods*, *5*, 362–387. https://doi.org/10.1177/109442802237116

Blumler, J. G. (1985). The social character of media gratifications. In K. E. Rosengren, L. A. Wenner, & P. Palmgreen (Eds.), *Media gratifications research: Current perspectives* (pp. 41–60). Beverly Hills, CA: Sage.

Boen, F., & Vanbeselaere, N. (2002). The impact of election outcome on the display of political posters: A field study during communal elections in Flanders. *Political Psychology*, *23*, 385–391. https://doi.org/10.1111/0162-895X.00286

Boen, F., Vanbeselaere, N., Pandelaere, M., Dewitte, S., Duriez, B., Snauwaert, B., … Avermaet, E. V. (2002). Politics and basking-in-reflected-glory: A field study in Flanders. *Basic and Applied Social Psychology*, *24*, 205–214. https://doi.org/10.1207/S15324834BASP2403_3

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York, NY: Guilford Press.

Boninger, D. S., Krosnick, J. A., & Berent, M. K. (1995). Origins of attitude importance: Self-interest, social identification, and value relevance. *Journal of Personality and Social Psychology*, *68*, 61. https://doi.org/10.1037/0022-3514.68.1.61

Brader, T., & Tucker, J. A. (2012). Following the party's lead: Party cues, policy opinion, and the power of partisanship in three multiparty systems. *Comparative Politics*, *44*, 403–415. https://doi.org/10.5129/001041512801283004

Branscombe, N. R., Ellemers, N., Spears, R., & Doosje, B. (1999). The context and content of social identity threat. In N. Ellemers, R. Spears, & B. Doosje (Eds.), *Social identity: Context, commitment, content* (pp. 35–58). Oxford, England: Blackwell Science.

Branscombe, N. R., & Wann, D. L. (1994). Collective self-esteem consequences of outgroup derogation when a valued social identity is on trial. *European Journal of Social Psychology*, *24*, 641–657. https://doi.org/10.1002/ejsp.2420240603

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin*, *17*, 475–482. https://doi.org/10.1177/0146167291175001

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., … Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLOS ONE*, *8*, e60188. https://doi.org/10.1371/journal.pone.0060188

Bürkner, P.-C. (2018). Advanced bayesian multilevel modeling with the R package brms. *The R Journal*, *10*, 395–411. https://doi.org/10.32614/RJ-2018-017

Cameron, J. E. (2004). A three-factor model of social identity. *Self and Identity*, *3*, 239–262. https://doi.org/10.1080/13576500444000047

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1980). *The American voter*. Chicago, IL: University of Chicago Press. (Original work published 1960)

Campbell, J. E. (1992). Forecasting the presidential vote in the states. *American Journal of Political Science*, *36*, 386–407. https://doi.org/10.2307/2111483

Cantin, S., & Boivin, M. (2004). Change and stability in children's social network and self-perceptions during transition from elementary to junior high school. *International Journal of Behavioral Development*, *28*, 561–570. https://doi.org/10.1080/01650250444000289

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). *Stan*: A probabilistic programming language. *Journal of Statistical Software*, *76*. https://doi.org/10.18637/jss.v076.i01

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*, 752–766.

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). London; Thousand Oaks, CA: SAGE.

Chong, D., & Druckman, J. N. (2010). Dynamic public opinion: Communication effects over time. *American Political Science Review*, *104*, 663–680. https://doi.org/10.1017/S0003055410000493

Cialdini, R. B., Borden, R. J., Thorne, A., Walker, M. R., Freeman, S., & Sloan, L. R. (1976). Basking in reflected glory: Three (football) field studies. *Journal of Personality and Social Psychology*, *34*, 366–375. https://doi.org/10.1037/0022-3514.34.3.366

Clark, T. S., & Linzer, D. A. (2015). Should I use fixed or random effects? *Political Science Research and Methods*, *3*, 399–408. https://doi.org/10.1017/psrm.2014.32

Clarke, H. D., & McCutcheon, A. L. (2009). The dynamics of party identification reconsidered. *Public Opinion Quarterly*, *73*, 704–728. https://doi.org/10.1093/poq/nfp051

Clifford, S. (2016). Individual differences in group loyalty predict partisan strength. *Political Behavior*. https://doi.org/10.1007/s11109-016-9367-3

Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from Mechanical Turk valid for research on political ideology? *Research & Politics*, *2*, 2053168015622072. https://doi.org/10.1177/2053168015622072

Comello, M. L. G., & Farman, L. (2016). Identity as a moderator and mediator of communication effects: Evidence and implications for message design. *The Journal of Psychology*, *150*, 822–836. https://doi.org/10.1080/00223980.2016.1196160

Converse, P. E. (1962). Information flow and the stability of partisan attitudes. *Public Opinion Quarterly*, *26*, 578–599. https://doi.org/10.1086/267129

Converse, P. E. (1964). The nature of belief systems in mass publics. In D. E. Apter (Ed.), *Ideology and Discontent* (pp. 206–261). Retrieved from http://ftp.voteview.com/The_Nature_of_Belief_Systems_in_Mass_Publics_Converse_1964.pdf

Coppock, A. (2018). Generalizing from survey experiments conducted on mechanical turk: A replication approach. *Political Science Research and Methods*, 1–16. https://doi.org/10.1017/psrm.2018.10

Coppock, A., & McClellan, O. A. (2019). Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. *Research & Politics*, *6*, 205316801882217. https://doi.org/10.1177/2053168018822174

Cornwell, B., Schumm, L. P., Laumann, E. O., Kim, J., & Kim, Y.-J. (2014). Assessment of social network change in a national longitudinal survey. *The Journals of Gerontology: Series B*, *69*, S75–S82. https://doi.org/10.1093/geronb/gbu037

Cowan, S. K., & Baldassarri, D. (2018). “It could turn ugly”: Selective disclosure of attitudes in political discussion networks. *Social Networks*, *52*, 1–17. https://doi.org/10.1016/j.socnet.2017.04.002

Craig, M. A., & Richeson, J. A. (2014). On the precipice of a “majority-minority” America: Perceived status threat from the racial demographic shift affects white Americans’ political ideology. *Psychological Science*, *25*, 1189–1197. https://doi.org/10.1177/0956797614527113

Crocker, J., & Luhtanen, R. (1990). Collective self-esteem and ingroup bias. *Journal of Personality and Social Psychology*, *58*, 60–67. https://doi.org/10.1037/0022-3514.58.1.60

Crocker, J., Luhtanen, R., Blaine, B., & Broadnax, S. (1994). Collective self-esteem and psychological well-being among White, Black, and Asian college students. *Personality and Social Psychology Bulletin*, *20*, 503–513. https://doi.org/10.1177/0146167294205007

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*, 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

de Vreese, C. H., Boukes, M., Schuck, A., Vliegenthart, R., Bos, L., & Lelkes, Y. (2017). Linking survey and media content data: Opportunities, considerations, and pitfalls. *Communication Methods and Measures*, *11*, 221–244. https://doi.org/10.1080/19312458.2017.1380175

de Vreese, C. H., & Neijens, P. (2016). Measuring media exposure in a changing communications environment. *Communication Methods and Measures*, *10*, 69–80. https://doi.org/10.1080/19312458.2016.1150441

Dennis, E. E., & Wartella, E. A. (1996). *American communication research: The remembered history*. https://doi.org/10.4324/9780203811870

Dennis, J. (1988). Political independence in America, part I: On being an independent partisan supporter. *British Journal of Political Science*, *18*, 77. https://doi.org/10.1017/S0007123400004968

Dennis, J. (1992). Political independence in America, III: In search of closet partisans. *Political Behavior*, *14*, 261–296. https://doi.org/10.1007/BF00991981

Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological Methods*, *22*, 240–261. https://doi.org/10.1037/met0000065

Dilliplane, S. (2014). Activation, conversion, or reinforcement? The impact of partisan news exposure on vote choice. *American Journal of Political Science*, *58*, 79–94. https://doi.org/10.1111/ajps.12046

Dilliplane, S., Goldman, S. K., & Mutz, D. C. (2013). Televised exposure to politics: New measures for a fragmented media environment. *American Journal of Political Science*, *57*, 236–248. https://doi.org/10.1111/j.1540-5907.2012.00600.x

Druckman, J. N. (2012). The politics of motivation. *Critical Review*, *24*, 199–216. https://doi.org/10.1080/08913811.2012.711022

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society*, *21*, 729–745. https://doi.org/10.1080/1369118X.2018.1428656

Egan, P. J. (2018, October 31). *Identity as dependent variable: How Americans shift their identities to better align with their politics*. Retrieved from https://www.dropbox.com/s/a78c5wygoabljtd/egan.identity.as.dv.latest.pdf

Ellemers, N., Kortekaas, P., & Ouwerkerk, J. W. (1999). Self-categorisation, commitment to the group and group self-esteem as related but distinct aspects of social identity. *European Journal of Social Psychology*, *29*, 371–389.

Estabrook, R., Grimm, K. J., & Bowles, R. P. (2012). A Monte Carlo simulation study of the reliability of intraindividual variability. *Psychology and Aging*, *27*, 560–576. https://doi.org/10.1037/a0026669

Ethier, K. A., & Deaux, K. (1994). Negotiating social identity when contexts change: Maintaining identification and responding to threat. *Journal of Personality and Social Psychology*, *67*, 243–251.

Eveland, W. P., Appiah, O., & Beck, P. A. (2018). Americans are more exposed to difference than we think: Capturing hidden exposure to political and racial difference. *Social Networks*, *52*, 192–200. https://doi.org/10.1016/j.socnet.2017.08.002

Eveland, W. P., Jr., Hutchens, M. J., & Shen, F. (2009). Exposure, attention, or "use" of news? Assessing aspects of the reliability and validity of a central concept in political communication research. *Communication Methods and Measures*, *3*, 223–244. https://doi.org/10.1080/19312450903378925

Eveland, W. P., Jr., & Kleinman, S. B. (2013). Comparing general and political discussion networks within voluntary organizations using social network analysis. *Political Behavior*, *35*, 65–87. https://doi.org/10.1007/s11109-011-9187-4

Fazio, R. H., Powell, M. C., & Williams, C. J. (1989). The role of attitude accessibility in the attitude-to-behavior process. *Journal of Consumer Research*, *16*, 280–288. https://doi.org/10.1086/209214

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–140. https://doi.org/10.1177/001872675400700202

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, and Company.

Finkel, S. E. (1993). Reexamining the "minimal effects" model in recent presidential campaigns. *The Journal of Politics*, *55*, 1–21. https://doi.org/10.2307/2132225

Fitzgerald, J. (2013). What does "political" mean to you? *Political Behavior*, *35*, 453–479. https://doi.org/10.1007/s11109-012-9212-2

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, *80*, 298–320. https://doi.org/10.1093/poq/nfw006

Fletcher, R., & Jenkins, J. (2019). *Polarisation and the news media in Europe: A literature review of the effect of news use on polarisation across Europe*. Retrieved from http://www.europarl.europa.eu/RegData/etudes/STUD/2019/634413/EPRS_STU(2019)634413_EN.pdf

Frey, D. (1986). Recent research on selective exposure to information. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 19, pp. 41–80). https://doi.org/10.1016/S0065-2601(08)60212-9

Garrett, R. K. (2009). Politically motivated reinforcement seeking: Reframing the selective exposure debate. *Journal of Communication*, *59*, 676–699. https://doi.org/10.1111/j.1460-2466.2009.01452.x

Garrett, R. K., Carnahan, D., & Lynch, E. K. (2013). A turn toward avoidance? Selective exposure to online political information, 2004–2008. *Political Behavior*, *35*, 113–134. https://doi.org/10.1007/s11109-011-9185-6

Garrett, R. K., Gvirsman, S. D., Johnson, B. K., Tsfati, Y., Neo, R., & Dal, A. (2014). Implications of pro- and counterattitudinal information exposure for affective polarization. *Human Communication Research*, *40*, 309–332. https://doi.org/10.1111/hcre.12028

Garrett, R. K., Long, J. A., & Jeong, M. S. (2019). From partisan media to misperception: Affective polarization as mediator. *Journal of Communication*, *69*, 490–517. https://doi.org/10.1093/joc/jqz028

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (Third). Boca Raton: CRC Press.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge; New York: Cambridge University Press.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*, 1360–1383. https://doi.org/10.1214/08-AOAS191

Gelman, A., & King, G. (1993). Why are American presidential election campaign polls so variable when votes are so predictable? *British Journal of Political Science*, *23*, 409. https://doi.org/10.1017/S0007123400006682

Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, *126*, 1799–1839. https://doi.org/10.1093/qje/qjr044

Gerber, A. S., Gimpel, J. G., Green, D. P., & Shaw, D. R. (2011). How large and long-lasting are the persuasive effects of televised campaign ads? Results from a randomized field experiment. *American Political Science Review*, *105*, 135–150. https://doi.org/10.1017/S000305541000047X

Gerbner, G., & Gross, L. (1976). Living with television: The violence profile. *Journal of Communication*, *26*, 172–199. https://doi.org/10.1111/j.1460-2466.1976.tb01397.x

Gervais, B. T. (2014). Following the news? Reception of uncivil partisan media and the use of incivility in political expression. *Political Communication*, *31*, 564–583. https://doi.org/10.1080/10584609.2013.852640

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. *Contexts of Accommodation: Developments in Applied Sociolinguistics*, *1*, 1–68.

Giles, H., & Johnson, P. (1987). Ethnolinguistic identity theory: A social psychological approach to language maintenance. *International Journal of the Sociology of Language*, *1987*. https://doi.org/10.1515/ijsl.1987.68.69

Giner-Sorolla, R., & Chaiken, S. (1994). The causes of hostile media judgments. *Journal of Experimental Social Psychology*, *30*, 165–180. https://doi.org/10.1006/jesp.1994.1008

Gitlin, T. (1978). Media Sociology: The Dominant Paradigm. *Theory and Society*, *6*, 205–253. https://doi.org/10.2307/657009

Green, D. P., & Palmquist, B. (1990). Of artifacts and partisan instability. *American Journal of Political Science*, *34*, 872–902. https://doi.org/10.2307/2111402

Green, D. P., & Yoon, D. H. (2002). Reconciling individual and aggregate evidence concerning partisan stability: Applying time-series models to panel survey data. *Political Analysis*, *10*, 1–24. https://doi.org/10.1093/pan/10.1.1

Greene, S. (2002). The social-psychological measurement of partisanship. *Political Behavior*, *24*, 171–197. https://doi.org/10.1023/A:1021859907145

Greene, S. (2004). Social identity theory and party identification. *Social Science Quarterly*, *85*, 136–153.

Greenier, K. D., Kernis, M. H., McNamara, C. W., Waschull, S. B., Berry, A. J., Herlocker, C. E., & Abend, T. A. (1999). Individual differences in reactivity to daily events: Examining the roles of stability and level of self-esteem. *Journal of Personality*, *67*, 187–208. https://doi.org/10.1111/1467-6494.00052

Guess, A. M. (2015). Measure for measure: An experimental test of online political media exposure. *Political Analysis*, *23*, 59–75. https://doi.org/10.1093/pan/mpu010

Gunther, A. C., Miller, N., & Liebhart, J. L. (2009). Assimilation and contrast in a test of the hostile media effect. *Communication Research*, *36*, 747–764. https://doi.org/10.1177/0093650209346804

Hamaker, E. L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). New York, NY: Guilford.

Hamaker, E. L., Asparouhov, T., Brose, A., Schmiedek, F., & Muthén, B. (2018). At the frontiers of modeling intensive longitudinal data: Dynamic structural equation models for the affective measurements from the COGITO study. *Multivariate Behavioral Research*, *53*, 820–841. https://doi.org/10.1080/00273171.2018.1446819

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin*, *135*, 555–588. https://doi.org/10.1037/a0015701

Harwood, J. (1997). Viewing age: Lifespan identity and television viewing choices. *Journal of Broadcasting & Electronic Media*, *41*, 203–213. https://doi.org/10.1080/08838159709364401

Harwood, J. (1999). Age identification, social identity gratifications, and television viewing. *Journal of Broadcasting & Electronic Media*, *43*, 123–136. https://doi.org/10.1080/08838159909364479

Hastorf, A. H., & Cantril, H. (1954). They saw a game; a case study. *The Journal of Abnormal and Social Psychology*, *49*, 129–134. https://doi.org/10.1037/h0057880

Hayes, A. F., & Myers, T. A. (2009). Testing the "proximate casualties hypothesis": Local troop loss, attention to news, and support for military intervention. *Mass Communication and Society*, *12*, 379–402. https://doi.org/10.1080/15205430802484956

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, 0956797613480187.

Hecht, M. L. (1993). 2002—a research odyssey: Toward the development of a communication theory of identity. *Communication Monographs*, *60*, 76–82. https://doi.org/10.1080/03637759309376297

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2012). Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. *Statistics in Medicine*, *31*, 3328–3336. https://doi.org/10.1002/sim.5338

Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review*, *34*, 93–101. https://doi.org/10.2307/2092790

Heywood, H. B. (1931). On finite sequences of real numbers. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, *134*, 486–501. https://doi.org/10.1098/rspa.1931.0209

Hill, S. J., Lo, J., Vavreck, L., & Zaller, J. (2013). How quickly we forget: The duration of persuasion effects from mass communication. *Political Communication*, *30*, 521–547. https://doi.org/10.1080/10584609.2013.828143

Hoffman, L. (2007). Multilevel models for examining individual differences in within-person variation and covariation over time. *Multivariate Behavioral Research*, *42*, 609–629. https://doi.org/10.1080/00273170701710072

Hoffman, L. H., & Eveland, W. P., Jr. (2010). Assessing causality in the relationship between community attachment and local news media use. *Mass Communication and Society*, *13*, 174–195. https://doi.org/10.1080/15205430903012144

Holbert, R. L., Garrett, R. K., & Gleason, L. S. (2010). A new era of minimal effects? A response to Bennett and Iyengar. *Journal of Communication*, *60*, 15–34. https://doi.org/10.1111/j.1460-2466.2009.01470.x

Holbrook, T. (1996). *Do campaigns matter?* SAGE.

Honaker, J., & King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science*, *54*, 561–581. https://doi.org/10.1111/j.1540-5907.2010.00447.x

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, *45*. https://doi.org/10.18637/jss.v045.i07

Hornsey, M. J. (2008). Social identity theory and self-categorization theory: A historical review. *Social and Personality Psychology Compass*, *2*, 204–222. https://doi.org/10.1111/j.1751-9004.2007.00066.x

Hovland, C. I., Lumsdaine, A. A., & Sheffield, F. D. (1949). *Experiments on mass communication*. Princeton University Press.

Hox, J. J., & Stoel, R. D. (2005). Multilevel and SEM approaches to growth curve modeling. In B. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (Vol. 3, pp. 1296–1305). Hoboken, N.J: John Wiley & Sons.

Huckfeldt, R. R., Johnson, P. E., & Sprague, J. D. (2004). *Political disagreement: The survival of diverse opinions within communication networks*. Cambridge University Press.

Huddy, L. (2001). From social to political identity: A critical examination of social identity theory. *Political Psychology*, *22*, 127–156. https://doi.org/10.1111/0162-895X.00230

Huddy, L., Bankert, A., & Davies, C. (2018). Expressive versus instrumental partisanship in multiparty European systems. *Political Psychology*, *39*, 173–199. https://doi.org/10.1111/pops.12482

Huddy, L., Mason, L., & Aarøe, L. (2015). Expressive partisanship: Campaign involvement, political emotion, and partisan identity. *American Political Science Review*, *109*, 1–17. https://doi.org/10.1017/S0003055414000604

Hultsch, D. F., Strauss, E., Hunter, M. A., & MacDonald, S. W. S. (2008). Intraindividual variability, cognition, and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of cognition and aging* (1st ed., pp. 491–556). New York: Psychology Press.

Hunter, J. A., Cox, S. L., O'Brien, K., Stringer, M., Boyes, M., Banks, M., … Crawford, M. (2005). Threats to group value, domain-specific self-esteem and intergroup discrimination amongst minimal and national groups. *British Journal of Social Psychology*, *44*, 329–353. https://doi.org/10.1348/014466604X17939

Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, *59*, 19–39. https://doi.org/10.1111/j.1460-2466.2008.01402.x

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*, *76*, 405–431.

Jacobson, N. C. (2020). *Compliance thresholds in intensive longitudinal data: Worse than listwise deletion*. Presented at the Society for Ambulatory Assessment, Melbourne,

Australia. Retrieved from http://www.nicholasjacobson.com/talk/saa2020_compliance_thresholds/

Jacobson, N. C., Chow, S.-M., & Newman, M. G. (2019). The differential time-varying effect model (DTVEM): A tool for diagnosing and modeling time lags in intensive longitudinal data. *Behavior Research Methods*, *51*, 295–315. https://doi.org/10.3758/s13428-018-1101-0

Jagodzinski, W., & Kühnel, S. M. (1987). Estimation of reliability and stability in single-indicator multiple-wave models. *Sociological Methods & Research*, *15*, 219–258. https://doi.org/10.1177/0049124187015003003

Jennings, M. K., & Markus, G. B. (1984). Partisan orientations over the long haul: Results from the three-wave political socialization panel study. *The American Political Science Review*, *78*, 1000–1018. https://doi.org/10.2307/1955804

Ji, L., Chow, S.-M., Schermerhorn, A. C., Jacobson, N. C., & Cummings, E. M. (2018). Handling missing data in the modeling of intensive longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 715–736. https://doi.org/10.1080/10705511.2017.1417046

Johnston, R. (2006). Party identification: Unmoved mover or sum of preferences? *Annual Review of Political Science*, *9*, 329–351. https://doi.org/10.1146/annurev.polisci.9.062404.170523

Jongerling, J., Laurenceau, J.-P., & Hamaker, E. L. (2015). A multilevel AR(1) model: Allowing for inter-individual differences in trait-scores, inertia, and innovation variance.

Multivariate Behavioral Research, *50*, 334–349. https://doi.org/10.1080/00273171.2014.1003772

Jung, E., & Hecht, M. L. (2004). Elaborating the communication theory of identity: Identity gaps and communication outcomes. *Communication Quarterly*, *52*, 265–283. https://doi.org/10.1080/01463370409370197

Jürgens, P., Stark, B., & Magin, M. (2019). Two half-truths make a whole? On bias in self-reports and tracking data. *Social Science Computer Review*, 0894439319831643. https://doi.org/10.1177/0894439319831643

Kalmijn, M. (2012). Longitudinal analyses of the effects of age, marriage, and parenthood on social contacts and support. *Advances in Life Course Research*, *17*, 177–190. https://doi.org/10.1016/j.alcr.2012.08.002

Katz, E., Blumler, J. G., & Gurevitch, M. (1973). Uses and gratifications research. *Public Opinion Quarterly*, *37*, 509–523. https://doi.org/10.1086/268109

Katz, E., & Lazarsfeld, P. F. (1955). *Personal influence: The part played by people in the flow of mass communications*. Glencoe, Illinois: Free Press.

Keith, B. E., Magleby, D. B., Nelson, C. J., Orr, E., Westlye, M. C., & Wolfinger, R. E. (1992). *The myth of the independent voter*. Berkeley: University of California Press.

Kernis, M. H. (2005). Measuring self-esteem in context: The importance of stability of self-esteem in psychological functioning. *Journal of Personality*, *73*, 1569–1605. https://doi.org/10.1111/j.1467-6494.2005.00359.x

Kernis, M. H., Cornell, D. P., Sun, C.-R., Berry, A., & Harlow, T. (1993). There's more to self-esteem than whether it is high or low: The importance of stability of self-esteem. *Journal*

of Personality and Social Psychology, *65*, 1190–1204. https://doi.org/10.1037/0022-3514.65.6.1190

Klapper, J. T. (1960). *The effects of mass communication*. Free Press.

Klar, S. (2014). Identity and engagement among political independents in America. *Political Psychology*, *35*, 577–591. https://doi.org/10.1111/pops.12036

Klar, S., & Krupnikov, Y. (2016). *Independent politics: How american disdain for parties leads to political inaction*. https://doi.org/10.1017/CBO9781316471050

Klar, S., Krupnikov, Y., & Ryan, J. B. (2018). Affective polarization or partisan disdain? Untangling a dislike for the opposing party from a dislike of partisanship. *Public Opinion Quarterly*, *82*, 379–390. https://doi.org/10.1093/poq/nfy014

Klofstad, C. A., McClurg, S. D., & Rolfe, M. (2009). Measurement of political discussion networks: A comparison of two "name generator" procedures. *Public Opinion Quarterly*, *73*, 462–483. https://doi.org/10.1093/poq/nfp032

Klofstad, C. A., Sokhey, A. E., & McClurg, S. D. (2013). Disagreeing about disagreement: How conflict in social networks affects political behavior. *American Journal of Political Science*, *57*, 120–134. https://doi.org/10.1111/j.1540-5907.2012.00620.x

Knobloch-Westerwick, S., & Meng, J. (2009). Looking the other way: Selective exposure to attitude-consistent and counterattitudinal political information. *Communication Research*, *36*, 426–448. https://doi.org/10.1177/0093650209333030

Knobloch-Westerwick, S., & Meng, J. (2011). Reinforcement of the political self through selective exposure to political messages. *Journal of Communication*, *61*, 349–368. https://doi.org/10.1111/j.1460-2466.2011.01543.x

Knowles, J. E., Frederick, C., & Whitworth, A. (2019). merTools: Tools for analyzing mixed effect regression models (Version 0.5.0). Retrieved from https://CRAN.R-project.org/package=merTools

Kruikemeier, S., & Shehata, A. (2017). News media use and political engagement among adolescents: An analysis of virtuous circles using panel data. *Political Communication*, *34*, 221–242. https://doi.org/10.1080/10584609.2016.1174760

Krupnikov, Y., & Levine, A. S. (2014). Cross-sample comparisons and external validity. *Journal of Experimental Political Science*, *1*, 59–80. https://doi.org/10.1017/xps.2014.7

Kuhn, M. H. (1960). Self-attitudes by age, sex, and professional training. *The Sociological Quarterly*, *1*, 39–54. https://doi.org/10.1111/j.1533-8525.1960.tb01459.x

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480. https://doi.org/10.1037//0033-2909.108.3.480

Lachlan, K. A., & Levy, D. R. (2016). BIRGing, CORFing, and Twitter activity following a political referendum: Examining social media activity concerning the 2014 Scottish independence vote. *Communication Research Reports*, *33*, 217–222. https://doi.org/10.1080/08824096.2016.1186625

LaCour, M. J., & Vavreck, L. (2014). Improving media measurement: Evidence from the field. *Political Communication*, *31*, 408–420. https://doi.org/10.1080/10584609.2014.921258

Laenen, A., Alonso, A., Molenberghs, G., & Vangeneugden, T. (2009). A family of measures to evaluate scale reliability in a longitudinal setting. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*, 237–253. https://doi.org/10.1111/j.1467-985X.2008.00554.x

Lang, K., & Lang, G. E. (1953). The unique perspective of television and its effect: A pilot study. *American Sociological Review*, *18*, 3. https://doi.org/10.2307/2087842

Lang, K., & Lang, G. E. (2006). Personal influence and the new paradigm: Some inadvertent consequences. *The ANNALS of the American Academy of Political and Social Science*, *608*, 157–178. https://doi.org/10.1177/0002716206292614

Lazarsfeld, P. F. (1948). Communication research and the social psychologist. In W. Dennis (Ed.), *Current trends in social psychology* (pp. 218–273). Retrieved from http://archive.org/details/currenttrendsins00denn

Lazarsfeld, P. F., Berelson, B. R., & Gaudet, H. (1948). *The people's choice: How the voter makes up his mind in a presidential campaign* (2nd ed.). New York: Columbia University Press.

Lecheler, S., & de Vreese, C. H. (2011). Getting real: The duration of framing effects. *Journal of Communication*, *61*, 959–983. https://doi.org/10.1111/j.1460-2466.2011.01580.x

Lecheler, S., & de Vreese, C. H. (2013). What a difference a day makes? The effects of repetitive and competitive news framing over time. *Communication Research*, *40*, 147–175. https://doi.org/10.1177/0093650212470688

Lecheler, S., de Vreese, C., & Slothuus, R. (2009). Issue importance as a moderator of framing effects. *Communication Research*, *36*, 400–425. https://doi.org/10.1177/0093650209333028

Lee, C., Hornik, R., & Hennessy, M. (2008). The reliability and stability of general media exposure measures. *Communication Methods and Measures*, *2*, 6–22. https://doi.org/10.1080/19312450802063024

Leik, R. K., & Chalkley, M. A. (1997). On the stability of network relations under stress. *Social Networks*, *19*, 63–74. https://doi.org/10.1016/S0378-8733(96)00291-2

Leszczensky, L., & Wolbring, T. (2019). How to deal with reverse causality using panel data? Recommendations for researchers based on a simulation study. *Sociological Methods & Research*, *OnlineFirst*. https://doi.org/10.1177/0049124119882473

Lewis-Beck, M. S. (1985). Election forecasts in 1984: How accurate were they? *PS: Political Science & Politics*, *18*, 53–62. https://doi.org/10.1017/S1049096500021296

Livingstone, S. (1996). On the continuing problems of media effects research. In J. Curran & M. Gurevitch (Eds.), *Mass media and society* (2nd ed., pp. 305–324). London, UK: Edward Arnold.

Long, J. A. (2020). *panelr: Regression models and utilities for repeated measures and panel data [R Package]*. Retrieved from https://cran.r-project.org/package=panelr

Long, J. A., Eveland, W. P., Jr., & Slater, M. D. (2019). Partisan media selectivity and partisan identity threat: The role of social and geographic context. *Mass Communication and Society*, *22*, 145–170. https://doi.org/10.1080/15205436.2018.1504302

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley Pub. Co.

Lüdtke, O., Robitzsch, A., & Grund, S. (2017). Multiple imputation of missing data in multilevel designs: A comparison of different strategies. *Psychological Methods*, *22*, 141–165. https://doi.org/10.1037/met0000096

Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin*, *18*, 302–318. https://doi.org/10.1177/0146167292183006

Luong, K. T., & Knobloch-Westerwick, S. (2017). Can the media help women be better at math? Stereotype threat, selective exposure, media effects, and women's math performance. *Human Communication Research*, *43*, 193–213. https://doi.org/10.1111/hcre.12101

Mael, F. A., & Tetrick, L. E. (1992). Identifying organizational identification. *Educational and Psychological Measurement*, *52*, 813–824. https://doi.org/10.1177/0013164492052004002

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the bayesian framework. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02767

Marin, A. (2004). Are respondents more likely to list alters with certain characteristics? *Social Networks*, *26*, 289–307. https://doi.org/10.1016/j.socnet.2004.06.001

McGraw, K. M., Lodge, M., & Stroh, P. (1990). On-line processing in candidate evaluation: The effects of issue order, issue importance, and sophistication. *Political Behavior*, *12*, 41–58. https://doi.org/10.1007/BF00992331

McGuire, W. J. (1961). Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments. *The Journal of Abnormal and Social Psychology*, *63*, 326–332. https://doi.org/10.1037/h0048344

McLeod, J. M., Kosicki, G. M., & McLeod, D. M. (1994). The expanding boundaries of political communication effects. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (pp. 123–162). Hillsdale, NJ: Erlbaum.

Mead, G. H. (1934). *Mind, self, & society*. Chicago: University of Chicago Press.

Medeiros, M., & Noël, A. (2014). The forgotten side of partisanship: Negative party identification in four Anglo-American democracies. *Comparative Political Studies*, *47*, 1022–1046. https://doi.org/10.1177/0010414013488560

Menon, G. (1993). The effects of accessibility of information in memory on judgments of behavioral frequencies. *Journal of Consumer Research*, *20*, 431–440. https://doi.org/10.1086/209359

Menon, G. (1994). Judgments of behavioral frequencies: Memory search and retrieval strategies. In N. Schwarz & S. Sudman (Eds.), *Autobiographical Memory and the Validity of Retrospective Reports* (pp. 161–172). Retrieved from http://link.springer.com.proxy.lib.ohio-state.edu/chapter/10.1007/978-1-4612-2624-6_11

Menon, G. (1997). Are the parts better than the whole? The effects of decompositional questions on judgments of frequent behaviors. *Journal of Marketing Research*, *34*, 335–346. https://doi.org/10.2307/3151896

Miller, C. B. (2009). Yes we did! Basking in reflected glory and cutting off reflected failure in the 2008 presidential election. *Analyses of Social Issues and Public Policy*, *9*, 283–296. https://doi.org/10.1111/j.1530-2415.2009.01194.x

Minozzi, W., Song, H., Lazer, D. M. J., Neblo, M. A., & Ognyanova, K. (2020). The incidental pundit: Who talks politics with whom, and why? *American Journal of Political Science*, *64*, 135–151. https://doi.org/10.1111/ajps.12469

Moeller, J., Shehata, A., & Kruikemeier, S. (2018). Internet use and political interest: Growth curves, reinforcing spirals, and causal effects during adolescence. *Journal of Communication*, *68*, 1052–1078. https://doi.org/10.1093/joc/jqy062

Mollenhorst, G., Volker, B., & Flap, H. (2014). Changes in personal relationships: How social contexts affect the emergence and discontinuation of relationships. *Social Networks*, *37*, 65–80. https://doi.org/10.1016/j.socnet.2013.12.003

Morey, A. C., & Eveland, W. P., Jr. (2016). Measures of political talk frequency: Assessing reliability and meaning. *Communication Methods and Measures*, *10*, 51–68. https://doi.org/10.1080/19312458.2015.1118448

Morey, A. C., Eveland, W. P., Jr., & Hutchens, M. J. (2012). The "who" matters: Types of interpersonal relationships and avoidance of political disagreement. *Political Communication*, *29*, 86–103. https://doi.org/10.1080/10584609.2011.641070

Mroczek, D. K., & Spiro, A. (2003). Modeling intraindividual change in personality traits: Findings from the normative aging study. *The Journals of Gerontology: Series B*, *58*, P153–P165. https://doi.org/10.1093/geronb/58.3.P153

Mutz, D. C. (2006). *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University Press.

Mutz, D. C., & Reeves, B. (2005). The new videomalaise: Effects of televised incivility on political trust. *The American Political Science Review*, *99*, 1–15.

Nesselroade, J. R. (1991). The warp and woof of the developmental fabric. In R. Downs, L. Liben, & D. Palermo (Eds.), *Visions of development, the environment, and aesthetics: The legacy of Joachim F. Wohlwill* (pp. 213–240). Hillsdale, NJ: Erlbaum.

Nesselroade, J. R., & Ram, N. (2004). Studying intraindividual variability: What we have learned that will help us understand lives in context. *Research in Human Development*, *1*, 9–29. https://doi.org/10.1080/15427609.2004.9683328

Newman, D. A., & Sin, H.-P. (2009). How do missing data bias estimates of within-group agreement? Sensitivity of SD_{WG}, CV_{WG}, $r_{WG(J)}$, $r_{WG(J)}$*, and ICC to systematic nonresponse. *Organizational Research Methods*, *12*, 113–147. https://doi.org/10.1177/1094428106298969

Palmquist, B., & Green, D. P. (1992). Estimation of models with correlated measurement errors from panel data. *Sociological Methodology*, *22*, 119. https://doi.org/10.2307/270994

Petrocik, J. R. (1974). An analysis of intransitivities in the index of party identification. *Political Methodology*, *1*, 31–47.

Petrocik, J. R. (2009). Measuring party support: Leaners are not independents. *Electoral Studies*, *28*, 562–572. https://doi.org/10.1016/j.electstud.2009.05.022

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 19, pp. 123–205). https://doi.org/10.1016/S0065-2601(08)60214-2

Pfau, M., & Burgoon, M. (1988). Inoculation in political campaign communication. *Human Communication Research*, *15*, 91–111. https://doi.org/10.1111/j.1468-2958.1988.tb00172.x

Pooley, J. (2006). Fifteen pages that shook the field: *Personal Influence*, Edward Shils, and the remembered history of mass communication research. *The ANNALS of the American Academy of Political and Social Science*, *608*, 130–156. https://doi.org/10.1177/0002716206292460

Postmes, T., Spears, R., & Lea, M. (1998). Breaching or building social boundaries? SIDE-effects of computer-mediated communication. *Communication Research*, *25*, 689–715.

Potter, W. J. (2011). Conceptualizing mass media effect. *Journal of Communication*, *61*, 896–915. https://doi.org/10.1111/j.1460-2466.2011.01586.x

Prior, M. (2009). The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly*, *73*, 130–143. https://doi.org/10.1093/poq/nfp002

Prior, M. (2010). You've either got it or you don't? The stability of political interest over the life cycle. *The Journal of Politics*, *72*, 747–766. https://doi.org/10.1017/S0022381610000149

Prior, M. (2013a). Media and political polarization. *Annual Review of Political Science*, *16*, 101–127. https://doi.org/10.1146/annurev-polisci-100711-135242

Prior, M. (2013b). The challenge of measuring media exposure: Reply to Dilliplane, Goldman, and Mutz. *Political Communication*, *30*, 620–634. https://doi.org/10.1080/10584609.2013.819539

Ram, N., & Gerstorf, D. (2009). Time-structured and net intraindividual variability: Tools for examining the development of dynamic characteristics and processes. *Psychology and Aging*, *24*, 778–791. https://doi.org/10.1037/a0017915

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed). Thousand Oaks, CA: Sage.

Reid, S. A. (2012). A self-categorization explanation for the hostile media effect. *Journal of Communication*, *62*, 381–399. https://doi.org/10.1111/j.1460-2466.2012.01647.x

Reid, S. A., Giles, H., & Abrams, J. R. (2004). A social identity model of media usage and effects. *Zeitschrift Für Medienpsychologie*, *16*, 17–25. https://doi.org/10.1026/1617-6383.16.1.17

Roese, N. J., & Olson, J. M. (1994). Attitude importance as a function of repeated attitude expression. *Journal of Experimental Social Psychology*, *30*, 39–51. https://doi.org/10.1006/jesp.1994.1002

Rogelberg, S. G., Conway, J. M., Sederburg, M. E., Spitzmüller, C., Aziz, S., & Knight, W. E. (2003). Profiling active and passive nonrespondents to an organizational survey. *Journal of Applied Psychology*, *88*, 1104–1114. https://doi.org/10.1037/0021-9010.88.6.1104

Rogelberg, S. G., Luong, A., Sederburg, M. E., & Cristol, D. S. (2000). Employee attitude surveys: Examining the attitudes of noncompliant employees. *Journal of Applied Psychology*, *85*, 284–293. https://doi.org/10.1037/0021-9010.85.2.284

Rogers, C. R. (1959). A theory of therapy, personality, and interpersonal relationships, as developed in the client-centered framework. In S. Koch (Ed.), *Formulations of the person and the social context* (pp. 184–256). McGraw-Hill.

Rubin, M., & Hewstone, M. (1998). Social identity theory's self-esteem hypothesis: A review and some suggestions for clarification. *Personality and Social Psychology Review*, *2*, 40–62. https://doi.org/10.1207/s15327957pspr0201_3

Saleem, M., Wojcieszak, M. E., Hawkins, I., Li, M., & Ramasubramanian, S. (2019). Social identity threats: How media and discrimination affect Muslim Americans' identification

as Americans and trust in the U.S. government. *Journal of Communication*, *69*, 214–236. https://doi.org/10.1093/joc/jqz001

Scharkow, M. (2017). *Stable attitudes and behaviors as boundary conditions of reinforcing spiral models*. Presented at the 67th Annual International Communication Association Conference, San Diego, CA.

Scharkow, M. (2019). The reliability and temporal stability of self-reported media exposure: A meta-analysis. *Communication Methods and Measures*, *Advance online publication*, 1–14. https://doi.org/10.1080/19312458.2019.1594742

Scheepers, D., & Ellemers, N. (2005). When the pressure is up: The assessment of social identity threat in low and high status groups. *Journal of Experimental Social Psychology*, *41*, 192–200. https://doi.org/10.1016/j.jesp.2004.06.002

Schramm, W. (1959). Comments on "the state of communication research." *Public Opinion Quarterly*, *23*, 6–9. https://doi.org/10.1086/266841

Schultzberg, M., & Muthén, B. (2018). Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 495–515. https://doi.org/10.1080/10705511.2017.1392862

Schuurman, N. K., Ferrer, E., de Boer-Sonnenschein, M., & Hamaker, E. L. (2016). How to compare cross-lagged associations in a multilevel autoregressive model. *Psychological Methods*, *21*, 206–221. https://doi.org/10.1037/met0000062

Schwarz, N., Hippler, H.-J., Deutsch, B., & Strack, F. (1985). Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly*, *49*, 388–395. https://doi.org/10.1086/268936

Sears, D. O., & Freedman, J. L. (1967). Selective exposure to information: A critical review. *Public Opinion Quarterly*, *31*, 194–213.

Sears, D. O., Fu, M., Henry, P. J., & Bui, K. (2003). The origins and persistence of ethnic identity among the "new immigrant" groups. *Social Psychology Quarterly*, *66*, 419–437. https://doi.org/10.2307/1519838

Sellers, R. M., Smith, M. A., Shelton, J. N., Rowley, S. A. J., & Chavous, T. M. (1998). Multidimensional model of racial identity: A reconceptualization of African American racial identity. *Personality and Social Psychology Review*, *2*, 18–39. https://doi.org/10.1207/s15327957pspr0201_2

Shah, D. V., Cho, J., Eveland, W. P., & Kwak, N. (2005). Information and expression in a digital age: Modeling internet effects on civic participation. *Communication Research*, *32*, 531–565. https://doi.org/10.1177/0093650205279209

Shah, D. V., McLeod, D. M., Rojas, H., Cho, J., Wagner, M. W., & Friedland, L. A. (2017). Revising the communication mediation model for a new political communication ecology. *Human Communication Research*, *43*, 491–504. https://doi.org/10.1111/hcre.12115

Shehata, A., Ekström, M., & Olsson, T. (2016). Developing self-actualizing and dutiful citizens: Testing the AC-DC model using panel data among adolescents. *Communication Research*, *43*, 1141–1169.

Silvia, P. J., Kwapil, T. R., Walsh, M. A., & Myin-Germeys, I. (2014). Planned missing-data designs in experience-sampling research: Monte Carlo simulations of efficient designs for assessing within-person constructs. *Behavior Research Methods*, *46*, 41–54. https://doi.org/10.3758/s13428-013-0353-y

Slater, M. D. (2007). Reinforcing spirals: The mutual influence of media selectivity and media effects and their impact on individual behavior and social identity. *Communication Theory*, *17*, 281–303. https://doi.org/10.1111/j.1468-2885.2007.00296.x

Slater, M. D. (2015). Reinforcing spirals model: Conceptualizing the relationship between media content exposure and the development and maintenance of attitudes. *Media Psychology*, *18*, 370–395. https://doi.org/10.1080/15213269.2014.897236

Slater, M. D., & Hayes, A. F. (2010). The influence of youth music television viewership on changes in cigarette use and association with smoking peers: A social identity, reinforcing spirals perspective. *Communication Research*, *37*, 751–773.

Slater, M. D., Henry, K. L., Swaim, R. C., & Anderson, L. L. (2003). Violent media content and aggressiveness in adolescents: A downward spiral model. *Communication Research*, *30*, 713–736. https://doi.org/10.1177/0093650203258281

Slater, M. D., Henry, K. L., Swaim, R. C., & Cardador, J. M. (2004). Vulnerable teens, vulnerable times: How sensation seeking, alienation, and victimization moderate the violent media content-aggressiveness relation. *Communication Research*, *31*, 642–668. https://doi.org/10.1177/0093650204269265

Smith, B. K., Ehala, M., & Giles, H. (2017). Vitality theory. In *Oxford Research Encyclopedia of Communication*. https://doi.org/10.1093/acrefore/9780190228613.013.496

Snyder, C. R., Lassegard, M., & Ford, C. E. (1986). Distancing after group success and failure: Basking in reflected glory and cutting off reflected failure. *Journal of Personality and Social Psychology*, *51*, 382–388. https://doi.org/10.1037/0022-3514.51.2.382

Sokhey, A. E., Baker, A., & Djupe, P. A. (2015). The dynamics of socially supplied information: Examining discussion network stability over time. *International Journal of Public Opinion Research*, *27*, 565–587. https://doi.org/10.1093/ijpor/edv028

Sotirovic, M., & McLeod, J. M. (2001). Values, communication behavior, and political participation. *Political Communication*, *18*, 273–300. https://doi.org/10.1080/10584600152400347

Stanish, W. M., & Taylor, N. (1983). Estimation of the intraclass correlation coefficient for the analysis of covariance model. *The American Statistician*, *37*, 221–224. https://doi.org/10.2307/2683375

Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, *60*, 556–576. https://doi.org/10.1111/j.1460-2466.2010.01497.x

Stroud, N. J., Feldman, L., Wojcieszak, M., & Bimber, B. (2019). The consequences of forced versus selected political media exposure. *Human Communication Research*, *45*, 27–51. https://doi.org/10.1093/hcr/hqy012

Suárez-Álvarez, J., Pedrosa, I., & Lozano, L. M. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 149–158. https://doi.org/10.7334/psicothema2018.33

Suls, J. M., Green, P., & Hillis, S. (1998). Emotional reactivity to everyday problems, affective inertia, and neuroticism. *Personality and Social Psychology Bulletin*, *24*, 127–136. https://doi.org/10.1177/0146167298242002

Sunstein, C. R. (2001). *Republic.com*. Princeton, N.J.: Princeton University Press.

Sunstein, C. R. (2007). *Republic.com 2.0*. Princeton, N.J.; Oxford: Princeton University Press.

Swann, W. B., Jr. (1983). Self-verification: Bringing social reality into harmony with the self. In J. M. Suls & A. G. Greenwald (Eds.), *Social psychological perspectives on the self* (Vol. 2, pp. 33–66). Hillsdale, NJ: Lawrence Erlbaum.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*, 755–769.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, *223*, 96–102. https://doi.org/10.1038/scientificamerican1170-96

Tajfel, H. (1974). Social identity and intergroup behaviour. *Information (International Social Science Council)*, *13*, 65–93. https://doi.org/10.1177/053901847401300204

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worschel (Eds.), *The social psychology of intergroup relations* (pp. 33–48). Monterey, Calif.: Brooks/Cole.

Tanti, C., Stukas, A. A., Halloran, M. J., & Foddy, M. (2011). Social identity change: Shifts in social identity during adolescence. *Journal of Adolescence*, *34*, 555–567. https://doi.org/10.1016/j.adolescence.2010.05.012

Theodoridis, A. G. (2017). Me, myself, and (I), (D), or (R)? Partisanship and political cognition through the lens of implicit identity. *The Journal of Politics*, *79*, 1253–1267. https://doi.org/10.1086/692738

Trull, T. J., & Ebner-Priemer, U. W. (2020). Ambulatory assessment in psychopathology research: A review of recommended reporting guidelines and current practices. *Journal of Abnormal Psychology*, *129*, 56–63. https://doi.org/10.1037/abn0000473

Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2003). Stability of self-esteem across the life span. *Journal of Personality and Social Psychology*, *84*, 205–220. https://doi.org/10.1037/0022-3514.84.1.205

Tukey, J. W., & McLaughlin, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization 1. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, *25*, 331–352. Retrieved from JSTOR.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (Eds.). (1987). *Rediscovering the social group: A self-categorization theory*. Oxford: Basil Blackwell.

Turner, J. C., & Reynolds, K. J. (2012). Self-categorization theory. In P. Van Lange, A. Kruglanski, & E. Higgins, *Handbook of Theories of Social Psychology* (pp. 399–417). https://doi.org/10.4135/9781446249222.n46

Updegraff, J. A., Emanuel, A. S., Suh, E. M., & Gallagher, K. M. (2010). Sheltering the self from the storm: Self-construal abstractness and the stability of self-esteem. *Personality and Social Psychology Bulletin*, *36*, 97–108. https://doi.org/10.1177/0146167209353331

Valentino, N. A., Banks, A. J., Hutchings, V. L., & Davis, A. K. (2009). Selective exposure in the Internet age: The interaction between anxiety and information utility. *Political Psychology*, *30*, 591–613.

Vallone, R. P., Ross, L., & Lepper, M. R. (1985). The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of Personality and Social Psychology*, *49*, 577.

van Veelen, R., Otten, S., & Hansen, N. (2011). Linking self and ingroup: Self-anchoring as distinctive cognitive route to social identification. *European Journal of Social Psychology*, *41*, 628–637. https://doi.org/10.1002/ejsp.792

Wang, L. (Peggy), Hamaker, E., & Bergeman, C. S. (2012). Investigating inter-individual differences in short-term intra-individual variability. *Psychological Methods*, *17*, 567–581. https://doi.org/10.1037/a0029317

Wang, L. (Peggy), & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, *20*, 63–83. https://doi.org/10.1037/met0000030

Wang, Z., & Tchernev, J. M. (2012). The "myth" of media multitasking: Reciprocal dynamics of media multitasking, personal needs, and gratifications. *Journal of Communication*, *62*, 493–513. https://doi.org/10.1111/j.1460-2466.2012.01641.x

Wellman, B., Wong, R. Y., Tindall, D., & Nazer, N. (1997). A decade of network change: Turnover, persistence and stability in personal communities. *Social Networks*, *19*, 27–50. https://doi.org/10.1016/S0378-8733(96)00289-4

Wheaton, B., Muthén, B., Alwin, D. F., & Summers, G. F. (1977). Assessing reliability and stability in panel models. *Sociological Methodology*, *8*, 84–136. https://doi.org/10.2307/270754

Wilcox, R. (2005). Trimming and Winsorization. In *Encyclopedia of Biostatistics*. https://doi.org/10.1002/0470011815.b2a15165

Wiley, D. E., & Wiley, J. A. (1970). The estimation of measurement error in panel data. *American Sociological Review*, *35*, 112–117. https://doi.org/10.2307/2093858

Wiley, J. A., & Wiley, M. G. (1974). A note on correlated errors in repeated measurements. *Sociological Methods & Research*, *3*, 172–188. https://doi.org/10.1177/004912417400300202

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: SAGE.

Willis, G. B. (2008). Cognitive interviewing. In P. Lavrakas, *Encyclopedia of Survey Research Methods* (pp. 107–109). https://doi.org/10.4135/9781412963947.n73

Willson, S., & Miller, K. (2014). Data collection. In *Cognitive Interviewing Methodology* (pp. 15–33). https://doi.org/10.1002/9781118838860.ch3

Wojcieszak, M., Azrout, R., & de Vreese, C. (2017). Waving the red cloth: Media coverage of a contentious issue triggers polarization. *Public Opinion Quarterly*. https://doi.org/10.1093/poq/nfx040

Zhou, X., & Reiter, J. P. (2010). A note on Bayesian inference after multiple imputation. *The American Statistician*, *64*, 159–163. https://doi.org/10.1198/tast.2010.09109

Appendix A1. Questionnaire Measures

Political Discussion

In the first questionnaire:

1. In a typical week, how many days do you talk with someone about politics or news, either online or offline?
2. In the past month, about how many different people have you talked with about politics or news, online or offline?
3. Of those people, about what percentage do you think identify with or generally support…
    a. Republicans
    b. Democrats
    c. Another party or no party
4. **Yesterday,** how much did you talk (online or offline) about news or politics?
   [ ] hours [ ] minutes
   And about what percentage of that time was with people who…
    a. Are Republicans, support Republicans, or have a conservative point of view
       _____ %
    b. Are Democrats, support Democrats, or have a liberal point of view
       _____ %
    c. Do not support Republicans or Democrats and do not have a particularly conservative or liberal point of view
       _____ %

In daily questionnaires:

1. **Yesterday,** how much did you talk (online or offline) about news or politics?
   [ ] hours [ ] minutes
   And about what percentage of that time was with people who…
    a. Are Republicans, support Republicans, or have a conservative point of view
       _____ %
    b. Are Democrats, support Democrats, or have a liberal point of view
       _____ %
    c. Do not support Republicans or Democrats and do not have a particularly conservative or liberal point of view
       _____ %

Political Media

1. **Yesterday,** how much time did you spend reading, watching, listening to, or hearing about the news or political content, including posts you saw on social media?
   [ ] hours [ ] minutes

   And about what percentage of that time was the content from sources that…

   a. Sources that tend to favor the Republican party or conservative viewpoints. Examples of sources like this include *FOX News, Breitbart News, The Daily Wire/Ben Shapiro.*
      _____ %
   b. Sources that tend to favor the Democratic party or liberal viewpoints. Examples of sources like this include *MSNBC*, *Huffington Post, Mother Jones*.
      _____ %
   c. Sources that do not tend to favor one political party or ideology over another. Examples of sources like this include *USA Today, Politico, Yahoo! News*.
      _____ %

Political Identification

In the first questionnaire only:

1. Generally speaking, do you think of yourself as a…
    a. Republican
    b. Democrat
    c. Independent
    d. Something else
2. [If not a. or b. in Q1] Do you generally think of yourself as a little closer to the Republicans or Democrats?
    a. Closer to Republicans
    b. Closer to Democrats

On both the first and subsequent questionnaires:

3. Please rate your level of agreement with the following statements. Response choices are Strongly Disagree, Disagree, Disagree Somewhat, Agree Somewhat, Agree, Strongly Agree
    a. If I talk about [Republicans/Democrats] today, I would say "we" instead of "they."
    b. I am interested in what other people think about [Republicans/Democrats].
    c. If someone criticizes [Republicans/Democrats], it would feel like a personal insult.
    d. I have a lot in common with supporters of [Republicans/Democrats].
    e. If [Republicans/Democrats] do badly in a new opinion poll, it will ruin my day.
    f. If I meet someone who supports [Republicans/Democrats], I will feel connected with that person.
    g. If I talk about [Republicans/Democrats] today, I will refer to them as "my party."
    h. If someone praises [Republicans/Democrats], it will make me feel good.

Note: These items are based on the scale validation study by Bankert and colleagues (2017).

Collective Self-Esteem

1. Please rate your agreement with the following items. Response options are Strongly Disagree, Disagree, Disagree Somewhat, Agree Somewhat, Agree, Strongly Agree.
    a. Today, I regret that I am a [Republican/Democrat].*
    b. Today, I'm glad to be a [Republican/Democrat].
    c. I am feeling that being a [Republican/Democrat] is not worthwhile.*

d. I am feeling good about being a [Republican/Democrat].
e. I am happy that I am a [Republican/Democrat].
f. I feel that [Republicans/Democrats] have major accomplishments.
g. [Republicans/Democrats] are considered good by other people.
h. Most people consider [Republicans/Democrats] to be more ineffective than the [Democrats/Republicans] and other parties.*
i. Other people respect [Republicans/Democrats].
j. Other people think that [Republicans/Democrats] are unworthy.*
k. [Republicans/Democrats] are not respected by the broader society.*

Note: Items (a) through (f) tap the private subdimension of collective self-esteem. Asterisks indicate reverse-coded items. Wordings are adapted from Luhtanen and Crocker (1992) and Sellers, Smith, Shelton, Rowley, and Chavous (1998).

Other Items Included Only in Introductory Survey

1. How often do you pay attention to what's going on in government and politics?
    a. Always
    b. Most of the time
    c. About half the time
    d. Some of the time
    e. Never

2. In a typical week, how many days do you talk with someone about politics or news, either online or offline?

3. In the past month, about how many different people have you talked with about politics or news, online or offline?
4. Of those people, about what percentage do you think identify with or generally support…
    a. Republicans
    b. Democrats
    c. Another party or no party

5. We'd like to know a bit more about where you get your information about news and politics. Which of the following sources do you regularly use, including times you accessed these sources via social media? Please check any that you used **at least once in the past month**.
    a. ABC News
    b. Alternet
    c. BBC
    d. Breitbart
    e. Business Insider
    f. Buzzfeed
    g. CBS News
    h. CNN
    i. Columbus Dispatch
    j. Conservative Tribune
    k. Crooks and Liars
    l. Cybercast News Service (www.cnsnews.com)
    m. Daily Mail
    n. FactCheck (factcheck.org)
    o. FiveThirtyEight
    p. Fox News
    q. Full Frontal with Samantha Bee
    r. Huffington Post
    s. Infowars

t. LA Times
u. Last Week Tonight with John Oliver
v. Local radio stations
w. Local TV stations
x. Mother Jones
y. MoveOn.org
z. MSNBC
aa. National Review
bb. NBC News
cc. New York Post
dd. New York Times
ee. NewsBusters
ff. Newsmax
gg. Newsweek
hh. NPR
ii. NY Daily News
jj. One America News Network (OANN)
kk. Politico
ll. Politifact (politifact.com)
mm. Real Time with Bill Maher
nn. Right Wing News
oo. Salon
pp. Saturday Night Live
qq. Slate
rr. Snopes
ss. Talking Points Memo
tt. The Atlantic
uu. The Atlantic
vv. The Blaze
ww. The Daily Beast
xx. The Daily Caller
yy. The Daily Kos
zz. The Daily Show with Trevor Noah
aaa. The Daily Wire
bbb. The Democratic Underground
ccc. The Drudge Report
ddd. The Economist
eee. The Federalist
fff. The Guardian
ggg. The Intercept
hhh. The Nation
iii. The New Republic
jjj. The New Yorker
kkk. The Washington Times

lll. Time
mmm. USA Today
nnn. Vice
ooo. Vox
ppp. Wall Street Journal
qqq. Washington Post
rrr. Western Journalism
sss. Yahoo News
ttt. Young Conservatives

6. Some people think the government should provide fewer services even in areas such as health and education in order to reduce spending. Other people feel it is important for the government to provide many more services even if it means an increase in spending. And, of course, some other people have opinions somewhere in between. Where would you place yourself?
    a. 1 — government should provide many fewer services
    b. 2
    c. 3
    d. 4
    e. 5
    f. 6
    g. 7 — government should provide many more services
7. Some people believe that we should spend much less money for defense. Others feel that defense spending should be greatly increased. And, of course, some other people have opinions somewhere in between. Where would you place yourself?
    a. 1 — government should decrease defense spending
    b. 2
    c. 3
    d. 4
    e. 5
    f. 6
    g. 7 — government should increase defense spending
8. There is much concern about the rapid rise in medical and hospital costs. Some people feel there should be a government insurance plan which would cover all medical and hospital expenses for everyone. Others feel that all medical expenses should be paid by individuals through private insurance plans like Blue Cross or other company paid plans. And, of course, some other people have opinions somewhere in between. Where would you place yourself?
    a. 1 — everyone should be on government insurance that covers all costs
    b. 2
    c. 3
    d. 4
    e. 5

f. 6
g. 7 — everyone should be on private insurance and they/their employer covers costs

9. Some people feel the government should make every effort to improve the social and economic position of racial and ethnic minorities. Others feel that the government should not make any special efforts to help racial and ethnic minorities because they should help themselves. And, of course, some other people have opinions somewhere in between. Where would you place yourself?
    a. 1 — government should help racial and ethnic minorities
    b. 2
    c. 3
    d. 4
    e. 5
    f. 6
    g. 7 — racial and ethnic minorities should help themselves

10. We hear a lot of talk these days about liberals and conservatives. How would you describe yourself?
    a. Extremely liberal
    b. Liberal
    c. Slightly liberal
    d. Moderate; middle of the road
    e. Slightly conservative
    f. Conservative
    g. Extremely conservative

11. Generally speaking, do you think of yourself as a… [randomize order of Republican/Democrat]
    a. Republican
    b. Democrat
    c. Independent
    d. Something else
12. [If not a. or b. in previous question] Do you generally think of yourself as a little closer to the Republicans or Democrats?
    a. Closer to Republicans
    b. Closer to Democrats

13. [If previous question skipped] If you do not answer this question, you will not be able to continue in this study. You will still get credit for this survey, but you will not receive any of the follow-up surveys. Are you sure you are not closer to either the Democrats or Republicans?
    a. Yes, I'm sure and would like to exit the study now.

b. I am closer to Democrats.
c. I am closer to Republicans.

14. [If Democrat or closer to Democrats] If the Democratic Presidential primary elections were held today, which of the following candidates would you vote for? [randomize order of responses]
    a. Joe Biden
    b. Elizabeth Warren
    c. Bernie Sanders
    d. Pete Buttigieg
    e. Andrew Yang
    f. Someone else: _____
    g. I wouldn't support anyone running for the Democratic nomination

15. Are you…

    a. a man

    b. a woman

    c. Other: _____

16. Which of these categories describe you? Check all that apply.

    a. White/Caucasian

    b. Hispanic/Latino

    c. Black/African American

    d. Asian

    e. American Indian/Alaska Native

    f. Middle Eastern or North African

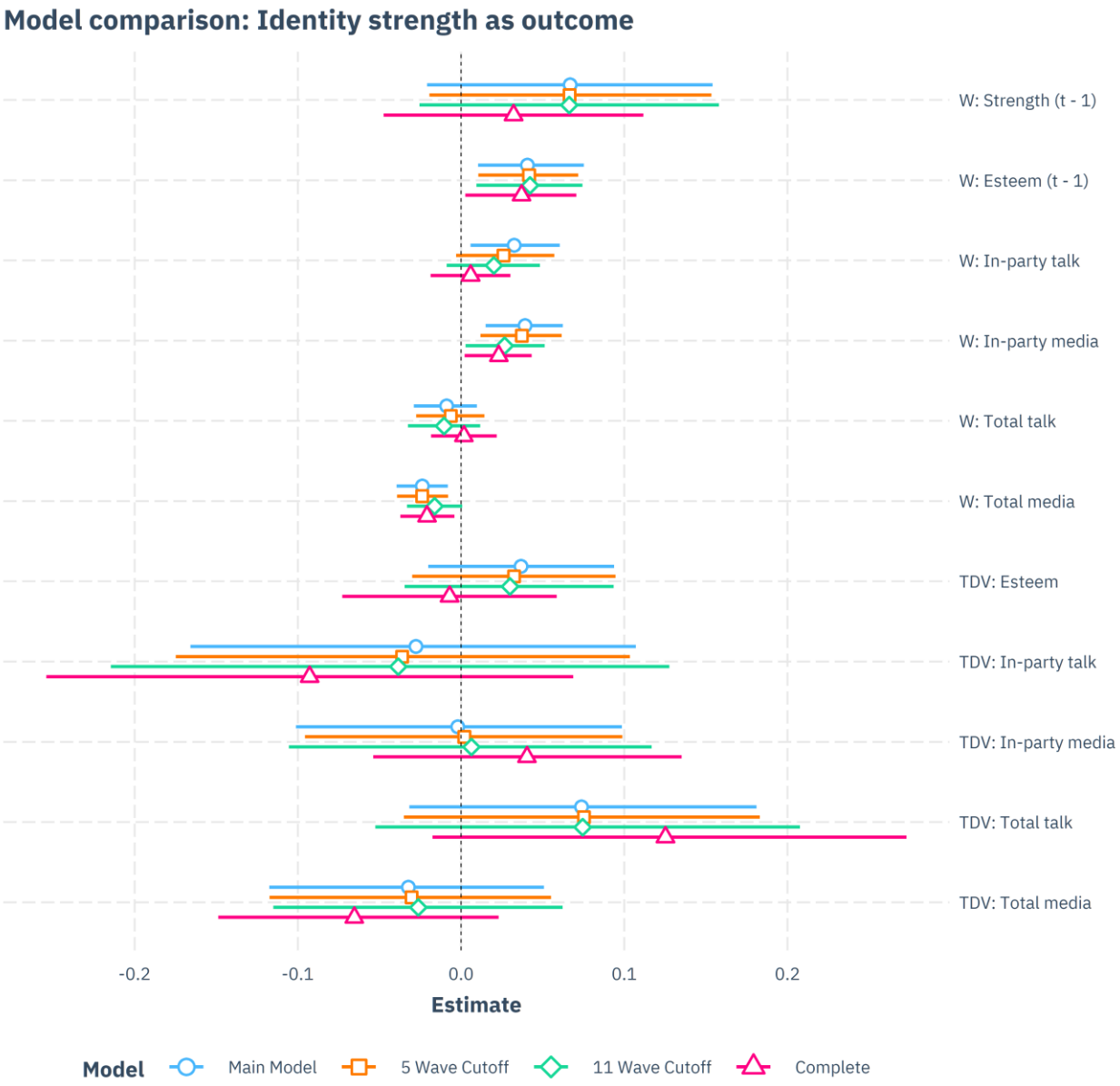
    g. Native Hawaiian/Pacific Islander

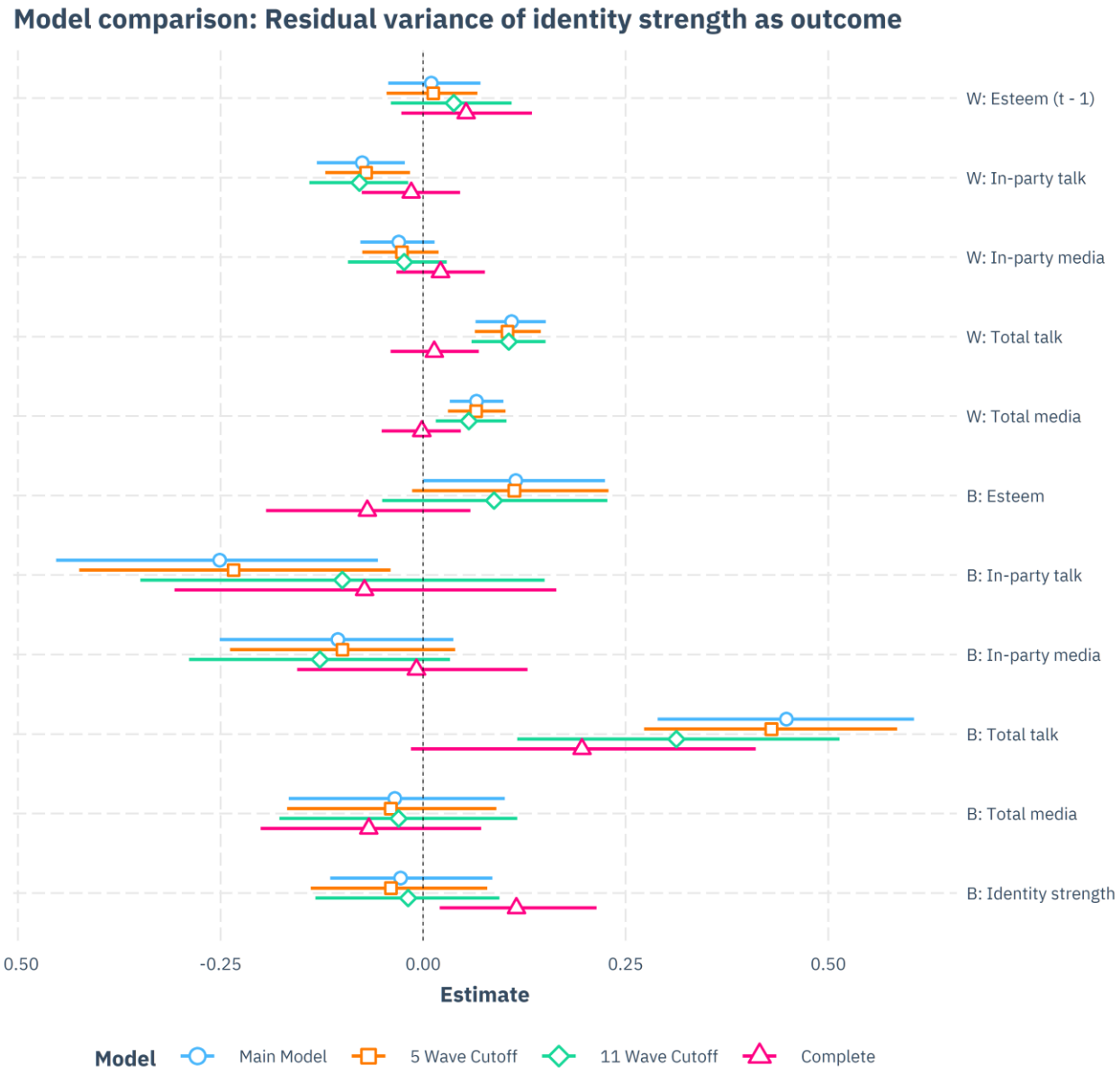
    h. Some other race, ethnicity, or origin

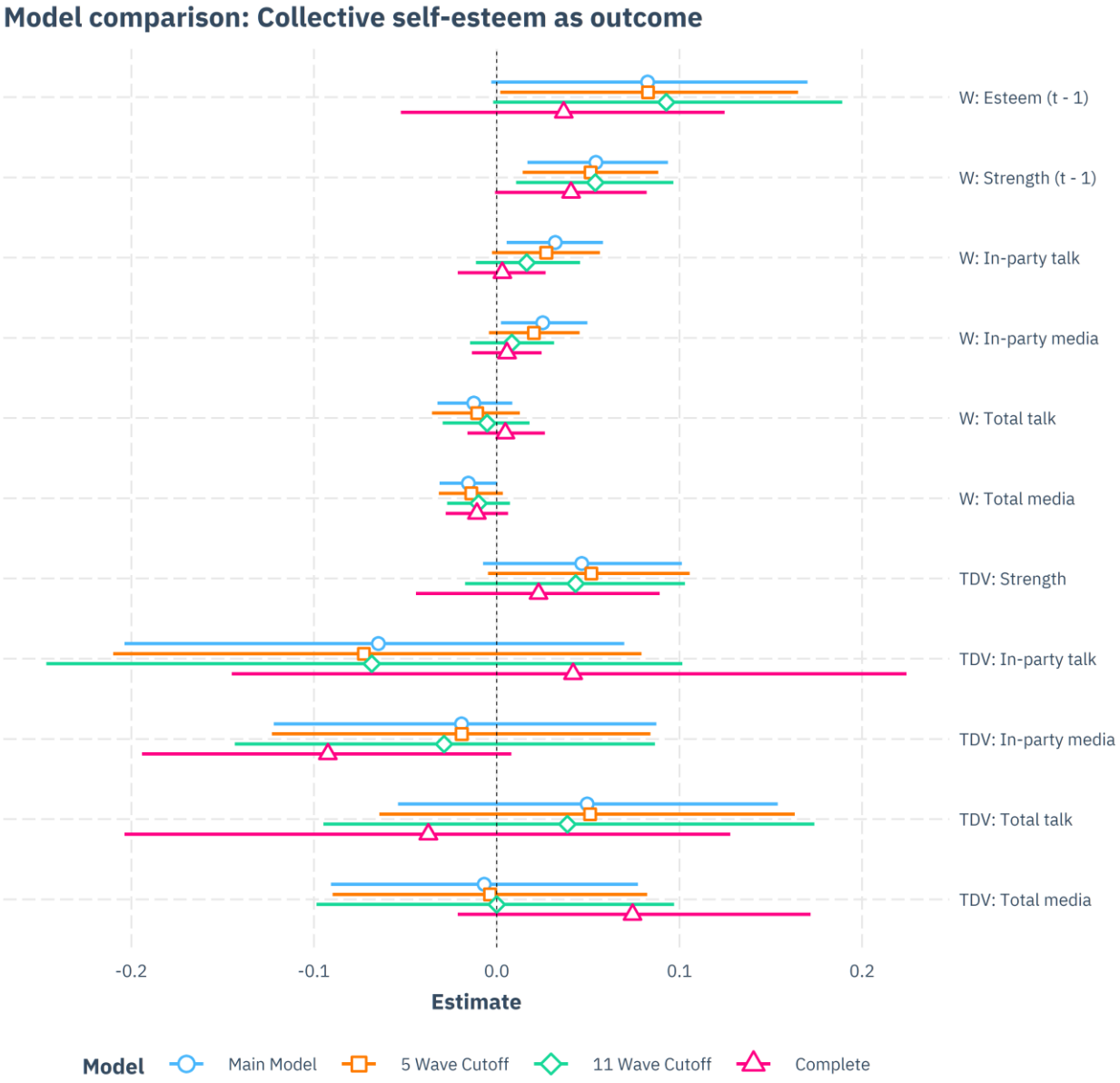
17. What is your year of birth?

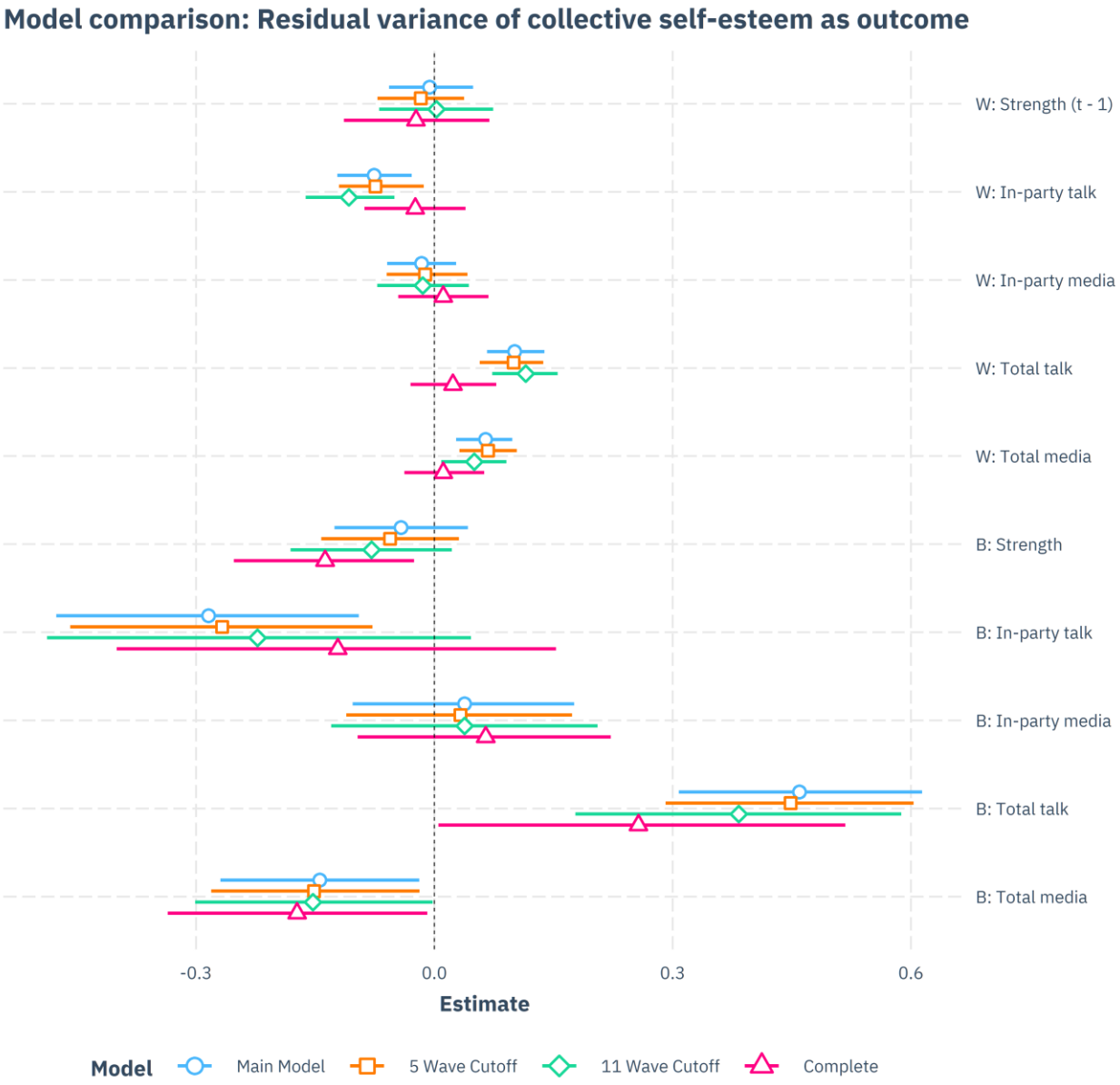
18. Have you ever voted in a U.S. election (at the local, state, and/or federal levels)? [Yes/No]

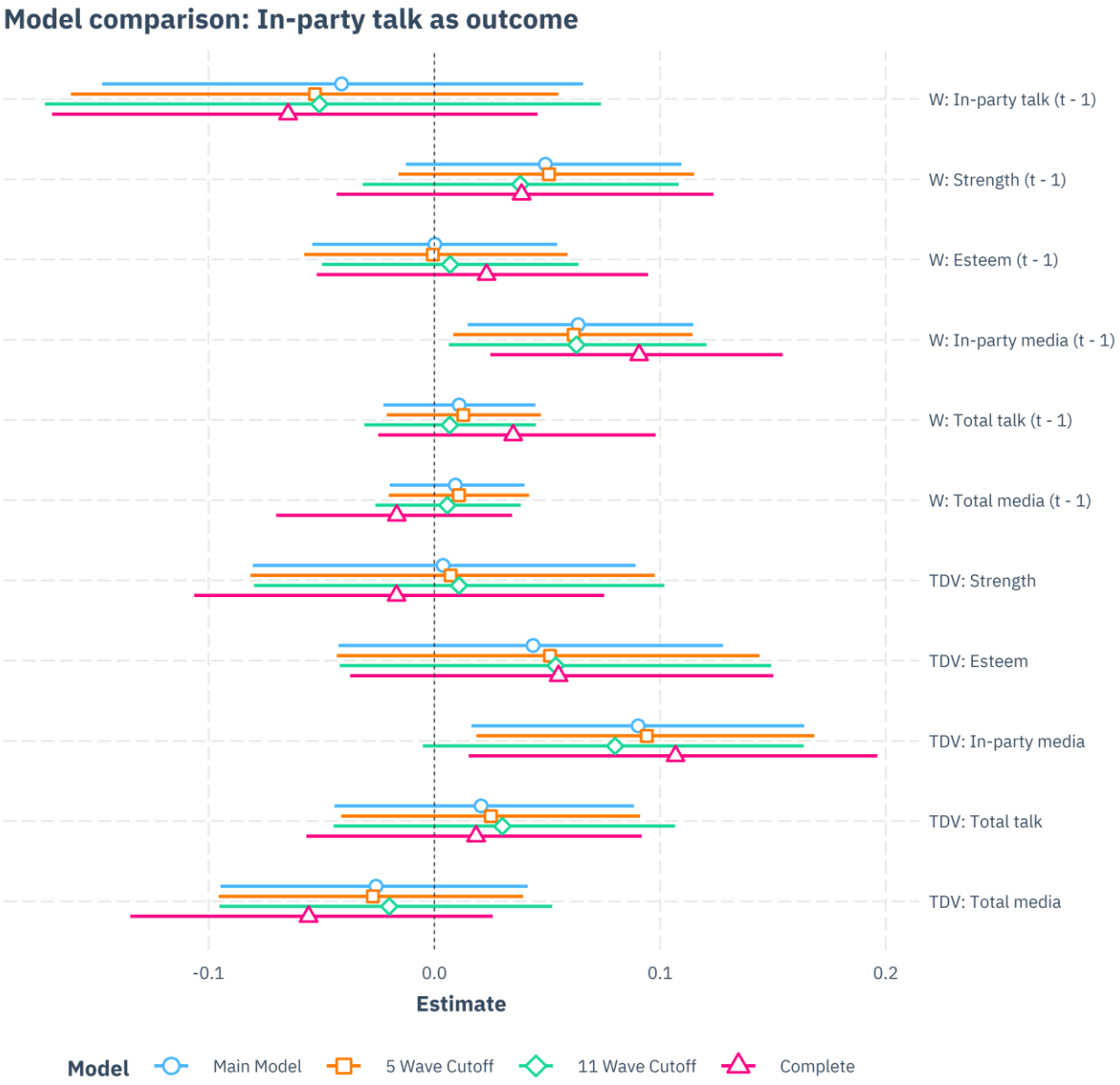
Appendix A2. Comparing Results from Multiple Imputation and Complete Data

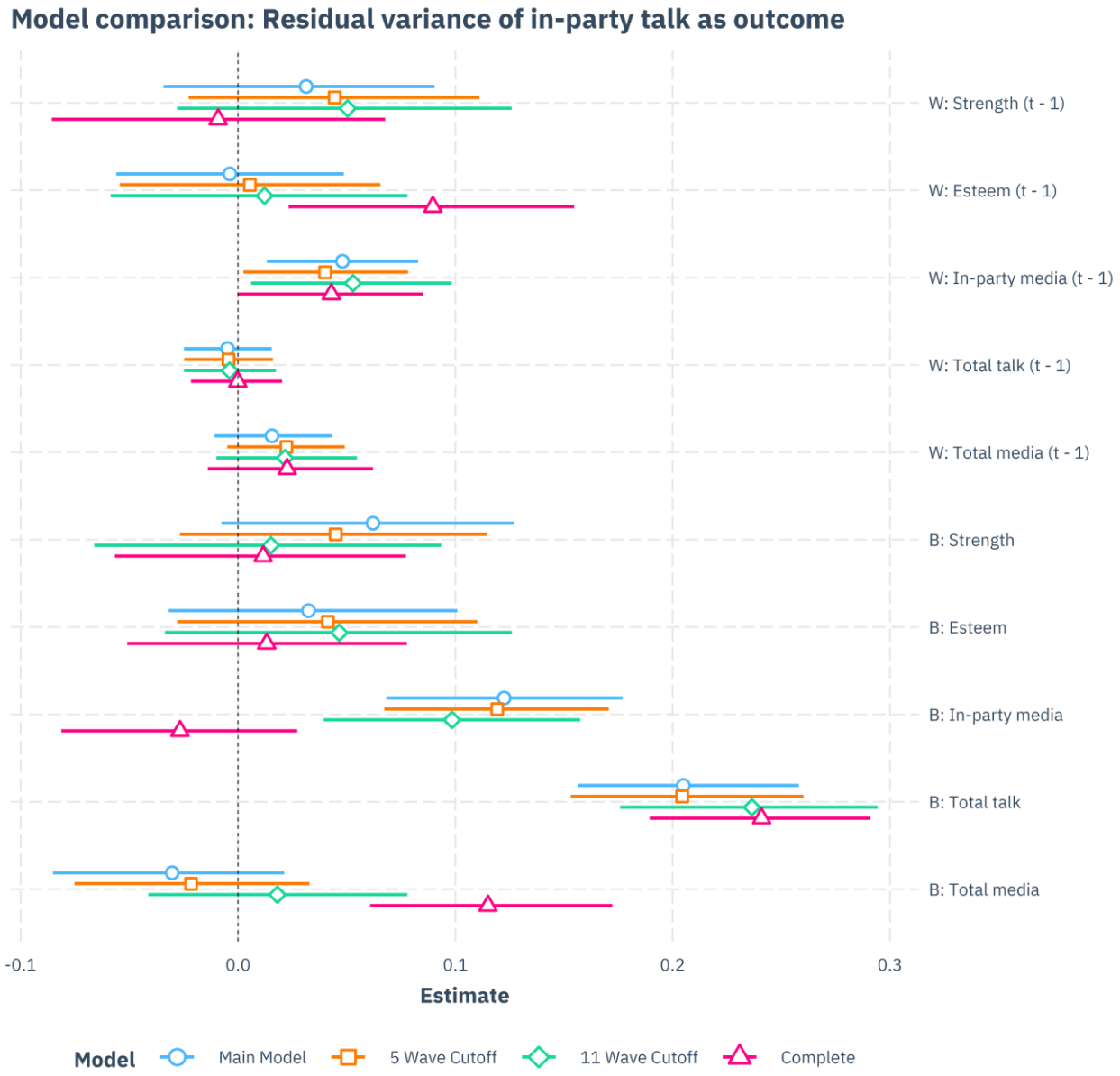


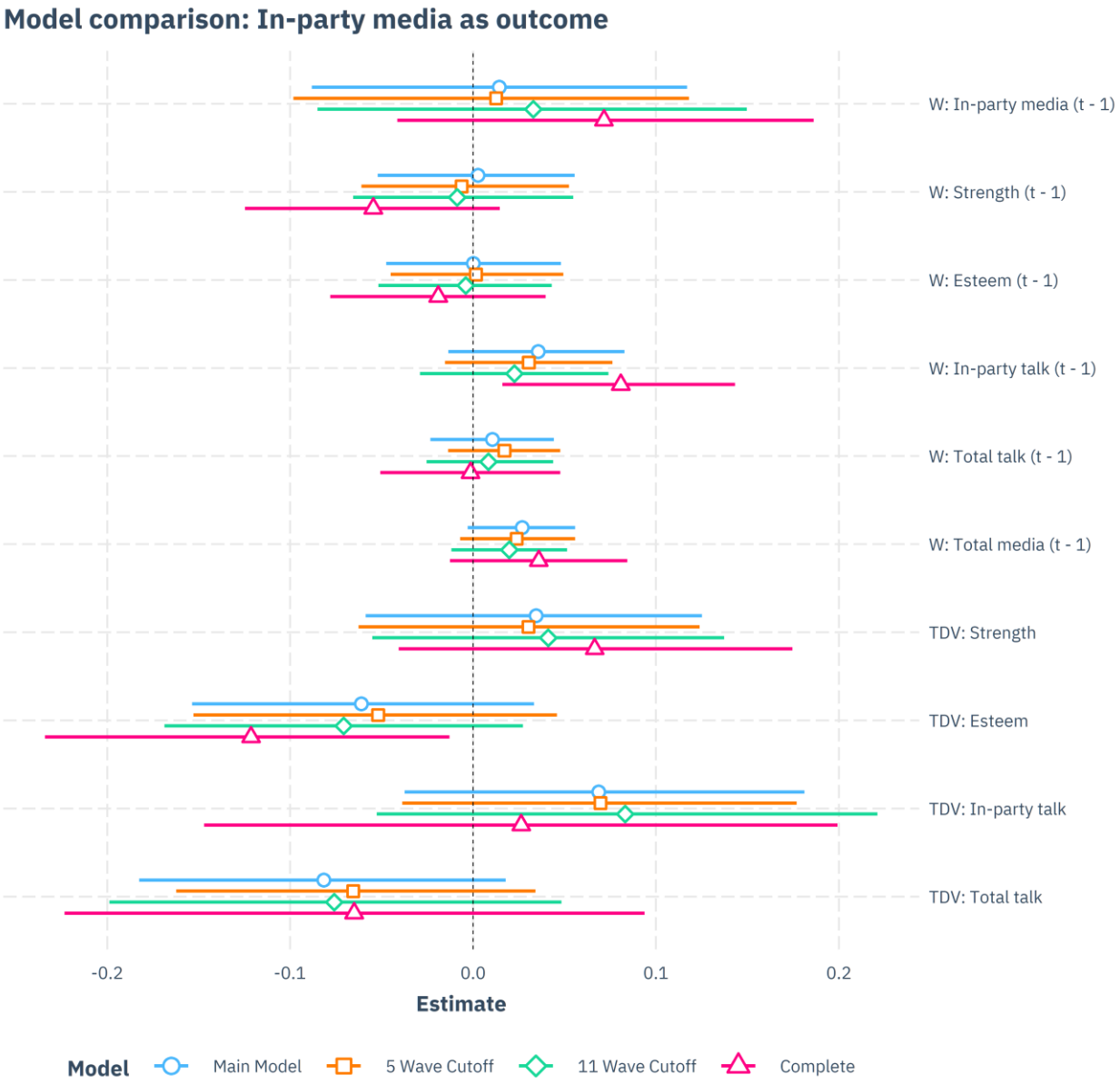

Model comparison: Residual variance of identity strength as outcome
W: Esteem (t - 1)
W: In-party talk
W: In-party media
W: Total talk
W: Total media
B: Esteem
B: In-party talk
B: In-party media
B: Total talk
B: Total media
B: Identity strength
0.50
-0.25
0.00
0.25
0.50
Estimate
Model
Main Model
5 Wave Cutoff
11 Wave Cutoff
Complete

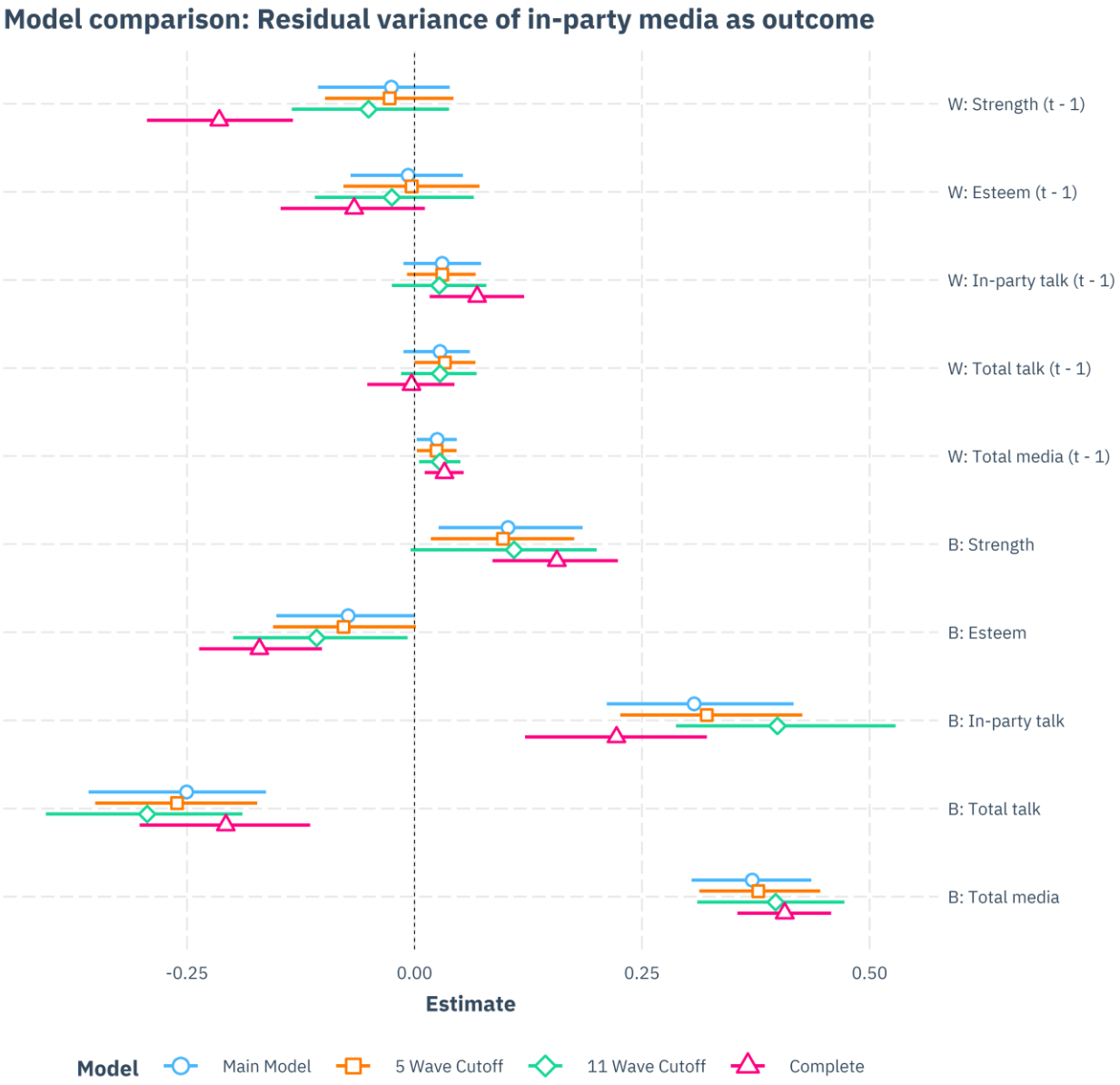

Model comparison: Collective self-esteem as outcome
W: Esteem (t - 1)
W: Strength (t - 1)
W: In-party talk
W: In-party media
W: Total talk
W: Total media
TDV: Strength
TDV: In-party talk
TDV: In-party media
TDV: Total talk
TDV: Total media
-0.2
-0.1
0.0
0.1
0.2
Estimate
Model
Main Model
5 Wave Cutoff
11 Wave Cutoff
Complete


Model comparison: Residual variance of collective self-esteem as outcome
W: Strength (t - 1)
W: In-party talk
W: In-party media
W: Total talk
W: Total media
B: Strength
B: In-party talk
B: In-party media
B: Total talk
B: Total media
-0.3
0.0
0.3
0.6
Estimate
Model
Main Model
5 Wave Cutoff
11 Wave Cutoff
Complete


Model comparison: In-party talk as outcome
W: In-party talk (t - 1)
W: Strength (t - 1)
W: Esteem (t - 1)
W: In-party media (t - 1)
W: Total talk (t - 1)
W: Total media (t - 1)
TDV: Strength
TDV: Esteem
TDV: In-party media
TDV: Total talk
TDV: Total media
-0.1
0.0
0.1
0.2
Estimate
Model
Main Model
5 Wave Cutoff
11 Wave Cutoff
Complete


Model comparison: Residual variance of in-party talk as outcome
W: Strength (t - 1)
W: Esteem (t - 1)
W: In-party media (t - 1)
W: Total talk (t - 1)
W: Total media (t - 1)
B: Strength
B: Esteem
B: In-party media
B: Total talk
B: Total media
-0.1
0.0
0.1
0.2
0.3
Estimate
Model
Main Model
5 Wave Cutoff
11 Wave Cutoff
Complete


Model comparison: In-party media as outcome
W: In-party media (t - 1)
W: Strength (t - 1)
W: Esteem (t - 1)
W: In-party talk (t - 1)
W: Total talk (t - 1)
W: Total media (t - 1)
TDV: Strength
TDV: Esteem
TDV: In-party talk
TDV: Total talk
-0.2
-0.1
0.0
0.1
0.2
Estimate
Model
Main Model
5 Wave Cutoff
11 Wave Cutoff
Complete


Model comparison: Residual variance of in-party media as outcome
W: Strength (t - 1)
W: Esteem (t - 1)
W: In-party talk (t - 1)
W: Total talk (t - 1)
W: Total media (t - 1)
B: Strength
B: Esteem
B: In-party talk
B: Total talk
B: Total media
-0.25
0.00
0.25
0.50
Estimate
Model
Main Model
5 Wave Cutoff
11 Wave Cutoff
Complete

Appendix A3. Missing Data Simulation Study

There are several analytical decisions, particularly regarding missing data, for which the right answer is not obvious in the case of intensive longitudinal designs. Although prior research indicates the importance of using imputation or similar procedures to account for missing observations (e.g., Ji et al., 2018), there are many details that are unclear, such as whether participants with few responses should be dropped from analysis (cf. Jacobson, 2020). And given that many missing data imputation methods — such as the one used in this study (Honaker & King, 2010) — incorporate both past and future observed values, it is not clear whether the analyst should impute values for dates beyond the last response of a given respondent. Although using future values to impute past values may seem counterintuitive from a causal inference point of view, Honaker and King write that "variables that are posttreatment, endogenously determined, or measures of the same quantity as others can all be helpful to include" and "including the dependent variable to impute missingness in an explanatory variable induces no endogeneity bias" (p. 563). One way to better understand the performance of different analytical decisions is to simulate data for which the true values are known and assess how each specification performs. This approach is generally called Monte Carlo simulation. What I did to assess these decisions is generate 500 simulated datasets reflecting several different possible ground truths about the process that generated the data collected in the main study for this dissertation.

Symbolically, the data are assumed to be generated by the following equations:

$$y_{it} = \alpha_y + \phi_y y_{it-1} + \beta_i x_{it-1} + \epsilon_{yit} \text{ with } \epsilon_{yit} \sim N(0, 1)$$

$$x_{it} = \alpha_x + \phi_x x_{it-1} + \epsilon_{xit} \text{ with } \epsilon_{xit} \sim N(0, 1)$$

$$\beta_i = \gamma_x + u_{xi} \text{ with } u_{xi} \sim N(0, \sigma_u)$$

That is, outcome variable y at time t and person i is determined by its own past value with autocorrelation ϕ_y (which for this analysis is arbitrarily set at 0.1), the past value of predictor variable x with coefficient β_i, and a random error that is normally distributed with mean 0 and standard deviation 1. The predictor x is, for the sake of simplicity, assumed to be caused by only its own past values with autocorrelation ϕ_x set at 0.1 and a normally distributed error with mean 0 and standard deviation of 1. As signaled by the subscript i, the data-generating process allows the effect of x on y to vary from one person to another with an average effect γ with a person-specific deviation from the average u_i. In the terminology of multilevel models, this is a varying slope.

Of particular interest is the effect of missing observations. Thus far, I have shown how the complete data are generated. One of the advantages of simulation is the complete data can be known before observations are marked as missing. The process of determining which observations to set to missing is important for exploring the effects of the available analytic decisions. To do this, each observation is assigned a probability of being completed, C_{it}, and the observation is determined complete or missing based on a single draw from a Bernoulli distribution with probability C_{it}. C_{it} is determined according to the following model:

$$logit(C_{it}) = \alpha_{Ci}$$

$$\alpha_{Ci} \sim N(\gamma_C, 1)$$

In other words, each person is assumed to have their own underlying propensity to complete surveys represented by α_{Ci}. The overall average γ_C is set such that at time 1, roughly 50% of respondents would be expected to complete the survey. This was chosen to approximate the observed missingness in this dissertation's main study.

Finally, although the previous explanations describe how the panel series are generated over time, one must also simulate the starting values; that is, the equations with a lagged value are undefined at the first wave. These are generated from a multivariate normal distribution that enables me to specify correlations between these values. To emulate the generally close cross-sectional relationship between media use and measures of political engagement, x and y are generated to have a correlation of 0.6 at time 1. Likewise, I can specify that some of these other parameters be correlated with one another. In all simulations, I specify that the individual-specific propensity to complete surveys α_{Ci} to have 0.3 correlation with both x and y to allow for the realistic possibility that missingness is related to the focal variables. These values make the data missing at random (MAR), meaning that the causes of missingness are systematic but related to observed variables and therefore feasibly accounted for via imputation methods and/or model specification. Another correlation is between the person-specific effect of x on y (β_i) and the person-specific propensity to complete surveys, α_{Ci}. In other words, this captures the possibility that people with greater values of β_i could be more or less likely to respond to surveys. Since the β_i is not directly observed, when there is a non-zero correlation between β_i and α_{Ci} the data are said to be missing not at random (MNAR). To summarize, the following is the correlation matrix of these starting values:

	x_0	y_0	α_{Ci}	β_i
x_0	1	—	—	—
y_0	0.6	1	—	—
α_{Ci}	0.3	0.3	1	—
β_i	0	0	[0.3, 0.1, 0, -0.1, -0.3]	1

Note that the correlation between β_i and α_{Ci} has multiple values because this is one of the key parameters whose influence is being tested in the simulations. Table 11 summarizes the parameters being varied in the simulations.

Table 11: Summary of Parameters Varied in Simulations

Parameter(s)	Substantive meaning	Values used
γ_x	Mean of β_i, effect of x on y	0; 0.25
$\rho(\beta_i, \alpha_{Ci})$	Correlation of person-specific effect of x on y and propensity to complete surveys	0.3; 0.1; 0; -0.1; -0.3

In sum, the simulations investigate the differences between several potential threats to the validity of estimates. One is whether the accuracy of the models differ depending on whether there is any effect of x to estimate (i.e., whether the average β_i is non-zero). Next is whether estimates depend on the extent to which effects vary by person. When this is set to 0, the effect of x is the same for every person; since some of the other varying parameters depend on covariance this one, conditions where effect variance is 0 and covariance with this parameter is not 0 are not included in the simulations. Last, I consider the possibility that the magnitude of the effect of x is directly related to the likelihood of survey non-response.

To put these parameters into layman's terms, it can be helpful to imagine that they correspond to the concepts under study. Suppose y is identity strength and x is in-party discussion. The estimate of β_i is the effect of in-party discussion on the mean level of identity strength. This simulation makes the plausible assumption that the effect of in-party discussion is not the same for everyone and that, perhaps, the effects of in-party discussion are weaker for those with greater identity strength. Although the simulation models this a bit differently, the potential problem being investigated is that the people who experience the largest effects from in-party discussion are the ones most likely to be infrequent survey responders. Varying the correlation between survey response tendency and the person-specific effect of x is meant to capture the consequences of this possibility.

The simulated datasets include 200 respondents for 21 potential surveys to approximate the dissertation study. I compare 5 analytical approaches: analyzing only the observed data, analyzing only the observed data for those who participate for at least half the surveys, doing

missing data imputation only for respondents who had at least 11 responses, doing missing data imputation only for those who had at least 3 responses, and doing missing data imputation only for those who had at least 3 responses and only imputing as far as the respondent's final response. Because the Bayesian method used for estimation in the main dissertation study is so computationally intense as to be time-prohibitive, the more straightforward maximum likelihood estimation method implemented in the R package lme4 (Bates, Mächler, Bolker, & Walker, 2015) is used for these purposes. Results from imputed datasets are averaged together using the R package merTools (Knowles, Frederick, & Whitworth, 2019). Each analysis uses 10 imputed datasets.

When evaluating the performance of each analytical decision, I use a few metrics: root mean squared error (RMSE), bias, and the false positive/negative rates of the estimates. For each of the 6 data scenarios, there are 500 simulated datasets and each method is used to analyze those same 500 datasets. The 500 estimates are extracted and compared to the known values of the parameters. For the purposes of this study, the focus is solely on the estimated β, the average effect of x on y. Generally speaking, no method performs best in every scenario, so one must evaluate the tradeoffs in terms of bias, variance, and so on when deciding the proper course of action given the results.

There are two basic categories of scenario to consider. One is in which x has no effect on y and therefore the interest is in the extent to which one might incorrectly conclude that there is one. The other kind is when x does affect y, in which case the interest is instead in one's ability to detect it. Figure 9 visualizes the most relevant consequences for the former scenario, in which

there is no effect and the most serious risk is of a false positive result. Here we see that a low threshold for inclusion is associated with a slightly inflated false positive rate relative to a higher threshold. Not doing imputation successfully avoids false positives in this scenario. Restricting imputation to measurements before the final observed wave differs only slightly from extrapolating beyond the final observed wave. Note that more computationally costly methods of calculating confidence intervals were not used in this simulation so it is possible that common methods for doing so would lower the false positive rate to its nominal value. A cruder method — simply using a 97.5% confidence interval and treating it as 95% confidence interval — approximately reduces the false positive rate to no more than 5% across all conditions. Regardless, the relative levels of false positivity are informative.

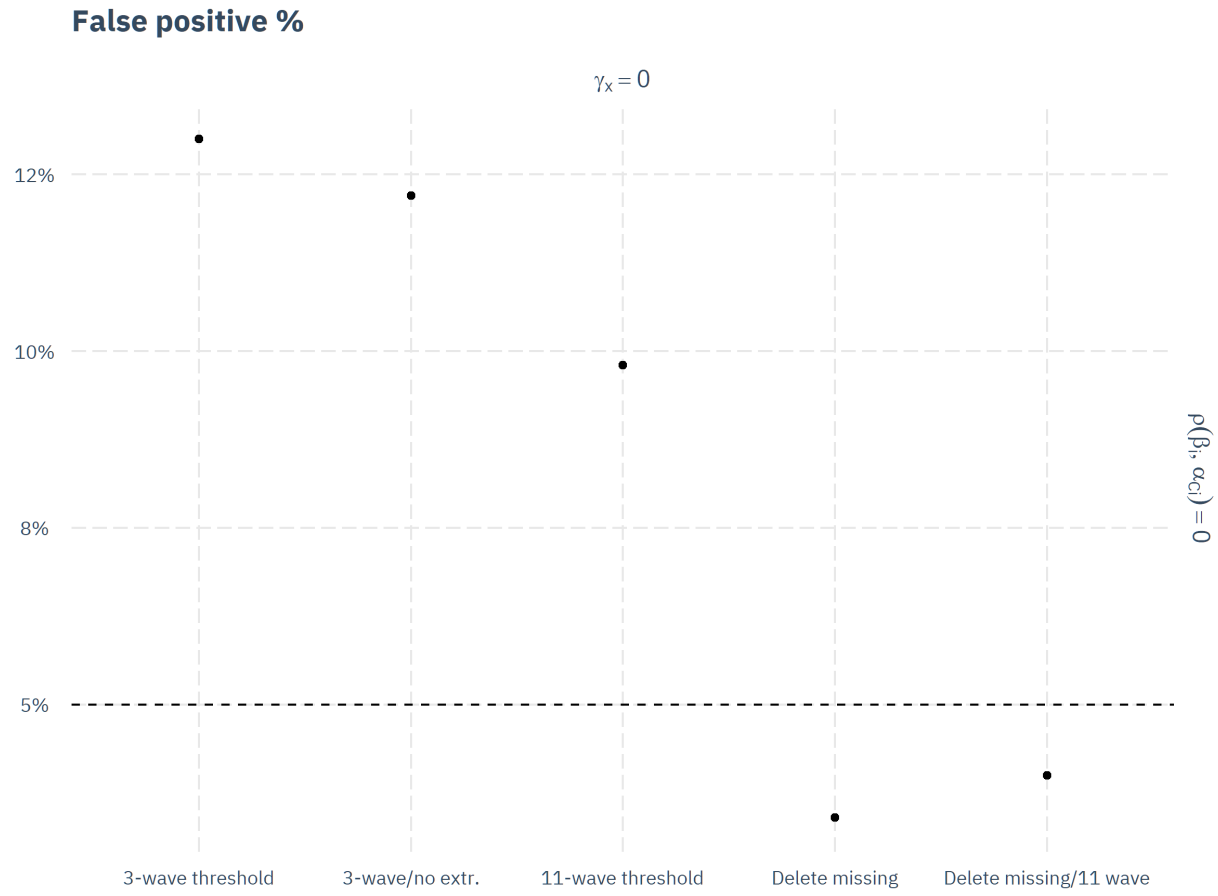


Figure 9: False Positive Rates Associated with Analytical Decisions

From here forward, I focus on the scenarios in which there is an effect of x on y. Figure 10 shows false negative rates across conditions. One key observation is that using a threshold for inclusion can impose a severe penalty to statistical power, even worse than keeping every observation and not doing any imputation, when data are MNAR. When data are MAR, there is still reduced statistical power, but this may be small enough to consider either complete case analysis or a threshold to be worth consideration given their other potential advantages. The situation in which this statistical power problem appears is that in which those who experience a greater effect of x on y are less likely to respond. The reason for this is reasonably straightforward: the available data do not evidence as large of an effect, biasing the estimates (see Figure 11) in the direction of the correlation. Note that this loss of power appears even with a fairly mild violation of the MAR assumption, a -0.1 correlation between the coefficient and the propensity to not respond. In this mild violation scenario, the true effect of x on y is positive for approximately 75% of participants so the problem is not merely that all the "responders" are excluded.

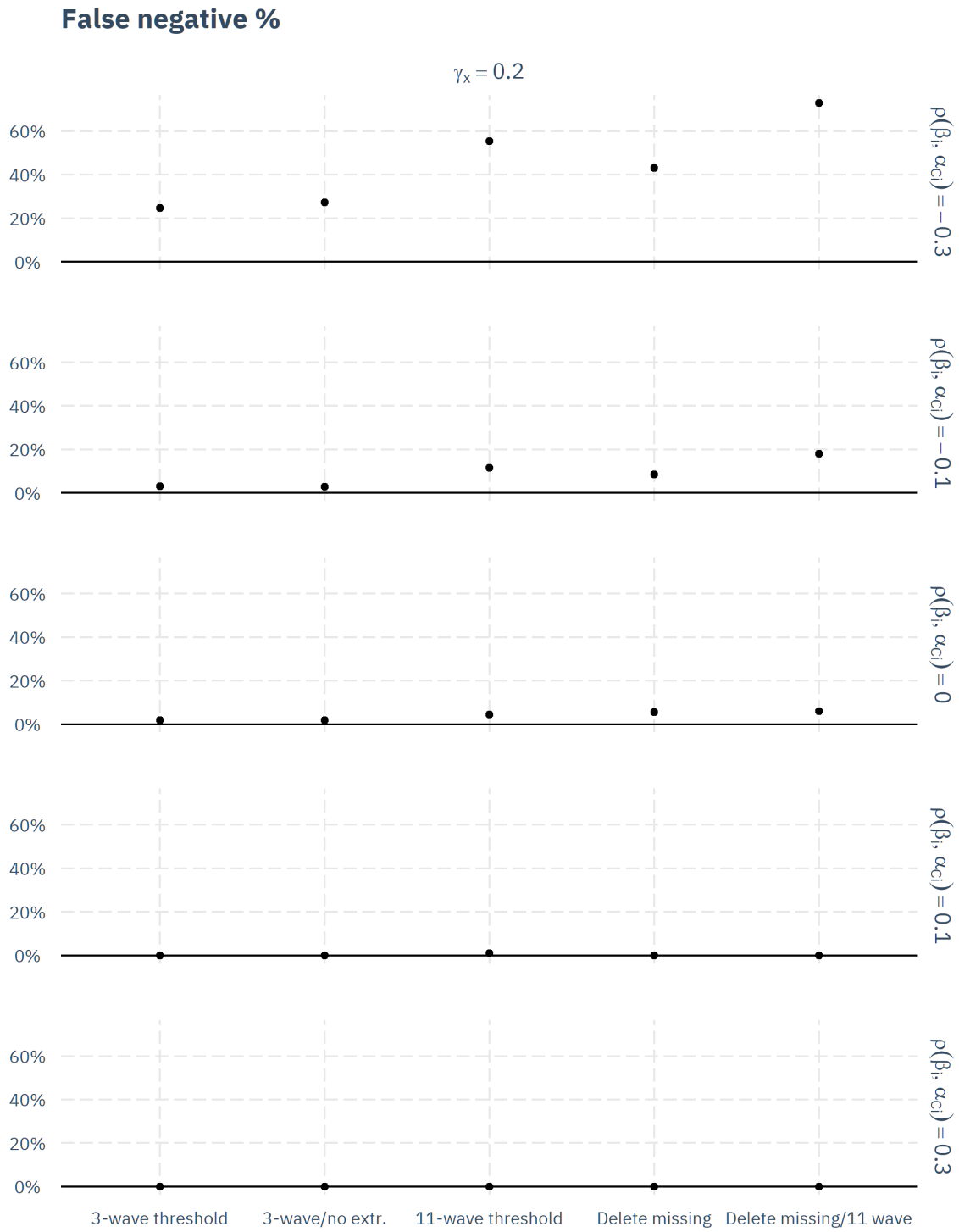


Figure 10: False Negative Rates Associated with Analytical Decisions

Another general pattern is that a lower threshold for inclusion imposes an attenuation bias, which one might expect to come at a cost to statistical power but in this case does not. In simulations not included here for simplicity, I have verified that this bias is not uniformly negative (i.e., if the effect of x on y is negative, the bias becomes positive, indicating the bias is towards 0 rather than consistently in one or the other direction). When it comes to RMSE (see Figure 12), there is no clearly superior option across all scenarios. Furthermore, in most scenarios there are no large differences among the specifications. The 3-wave threshold used in the main dissertation study often has a relatively high RMSE, but this can be explained by the aforementioned attenuation bias.

Overall, there's no unambiguous choice available. One thing that stands out is the performance of complete case analysis — although the goal was not exactly to stress-test this approach, it often out-performs multiple imputation with high participation thresholds. One reasonable takeaway from this analysis is that multiple imputation with a low threshold for inclusion offers the most statistical power with a modest drawback of potential for false positives when assumptions are not violated. In cases when a higher threshold performs better than a lower one, the option to use a lower threshold but not extrapolate beyond the final observation tends to somewhat split the difference but perform more similarly to full extrapolation. Results certainly give credence to Jacobson's (2020) advice to not impose thresholds for inclusion, although it is clear there are some cases in which it is less optimal; the challenge is foreseeing those scenarios. Another issue to consider is that I have focused only on main effects on the mean level of the dependent variable. It is possible that these findings could be different for predicting variance. Furthermore, in an effort to simplify the analysis, the simulation does not make any assumption

about whether people are more likely to not respond as the study goes on, even though there seems to be a clear pattern in the actual data in which responses become less likely over time. This analysis is not meant to be the final word on these issues, but rather to provide some preliminary data-based evidence on which to evaluate the available choices.

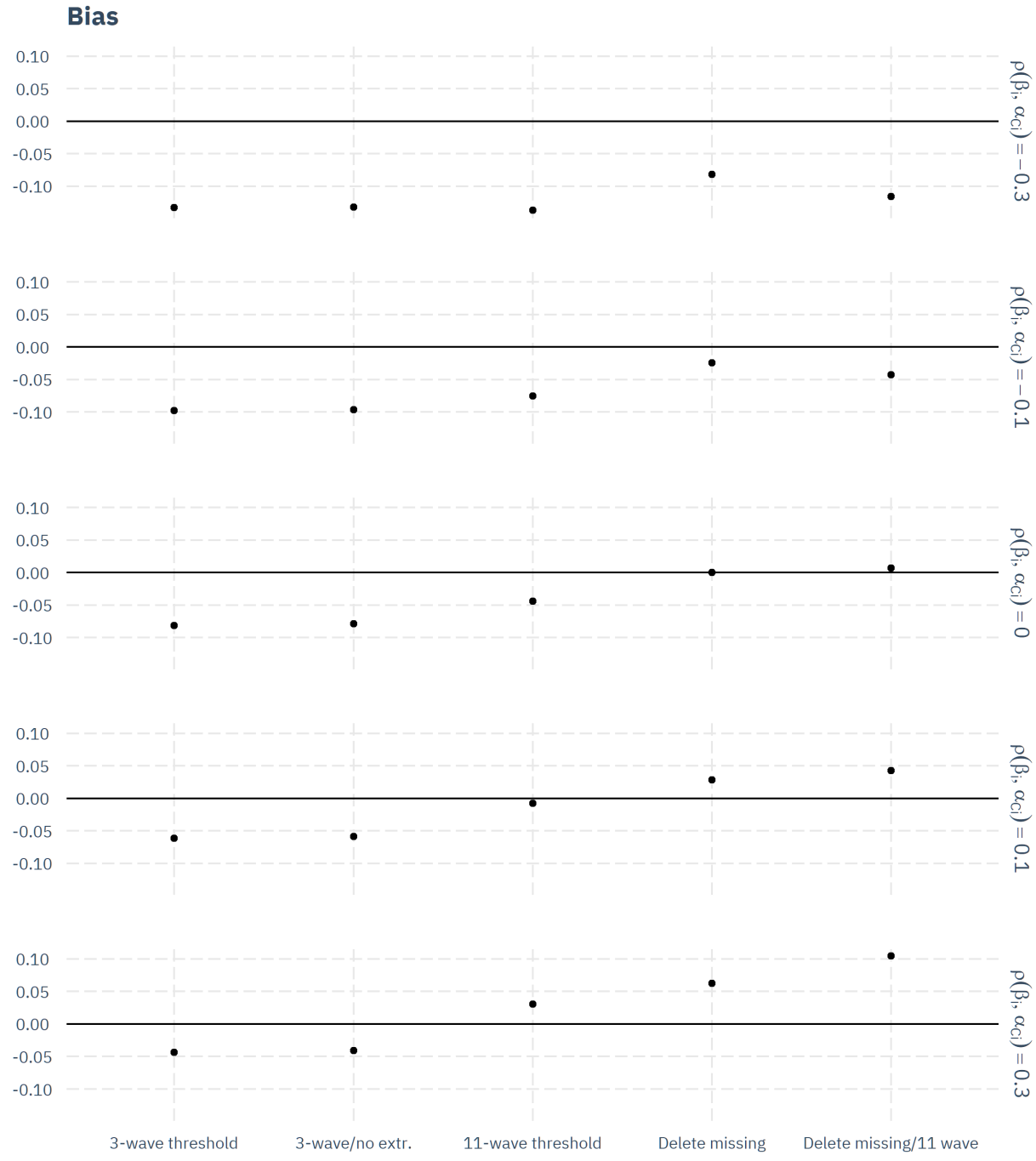


Figure 11: Bias Associated with Analytical Decisions

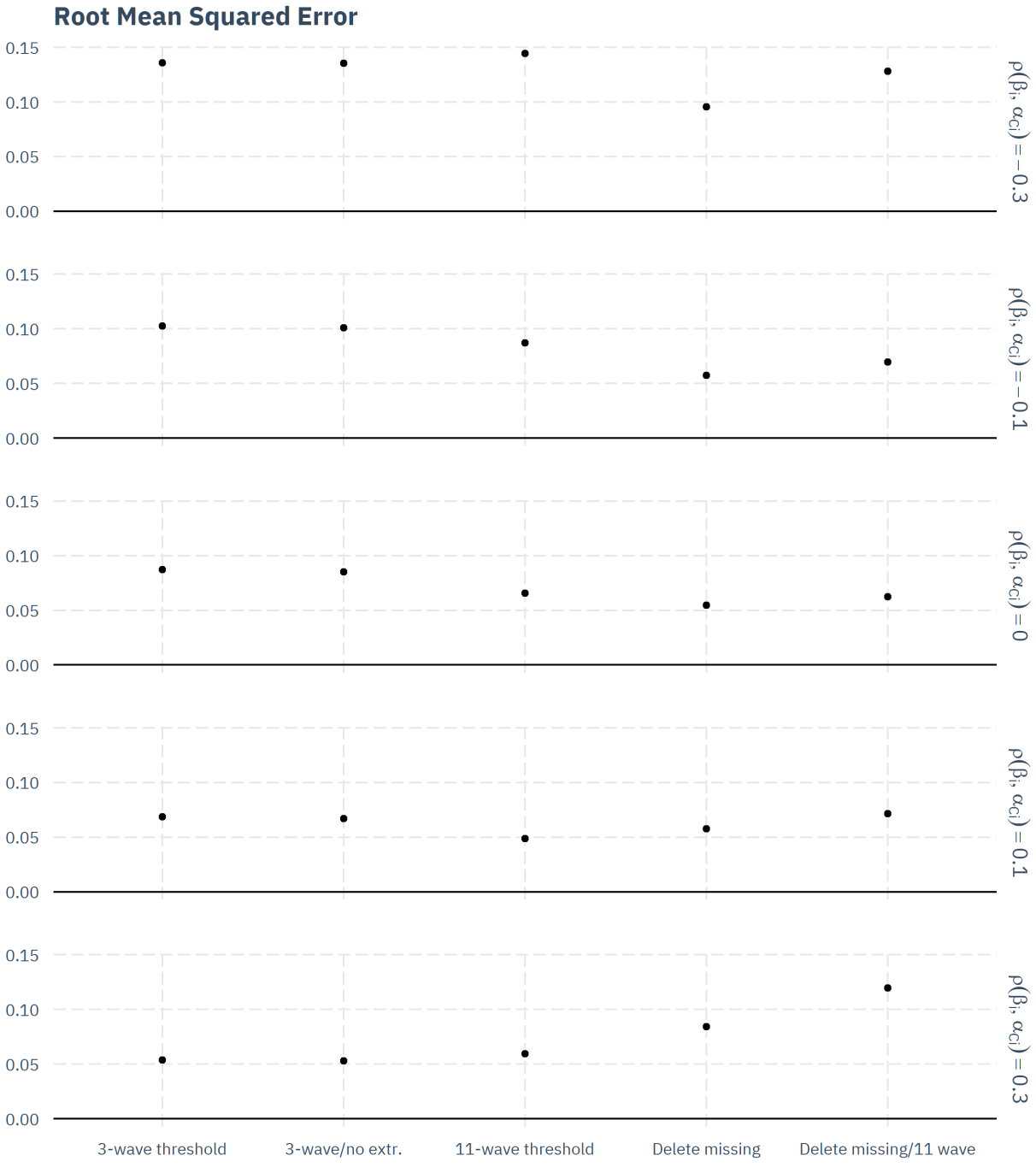


Figure 12: Estimation Error Associated with Analytical Decisions