Political psychophysiology

A primer for interested researchers and consumers

Jaime E. Settle, *College of William & Mary*

Matthew V. Hibbing, *University of California, Merced*

Nicolas M. Anspach, *York College of Pennsylvania*

Taylor N. Carlson, *Washington University in St. Louis*

Chelsea M. Coe, *University of California, Merced*

Edward Hernandez, *College of William & Mary*

John Peterson, *Palo Alto College*

John Stuart, *College of William & Mary*

Kevin Arceneaux, *Temple University*

ABSTRACT. **The past decade has seen a rapid increase in the number of studies employing psychophysiological methods to explain variation in political attitudes and behavior. However, the collection, analysis, and interpretation of physiological data present novel challenges for political scientists unfamiliar with the underlying biological concepts and technical skills necessary for utilizing this approach. Our objective in this article is to maximize the effectiveness of future work utilizing psychophysiological measurement by providing guidance on how the techniques can be employed most fruitfully as a complement to, not a replacement for, existing methods. We develop clear, step-by-step instructions for how physiological research should be conducted and provide a discussion of the issues commonly faced by scholars working with these measures. Our hope is that this article will be a useful resource for both neophytes and experienced scholars in lowering the start-up costs to doing this work and assessing it as part of the peer review process. More broadly, in the spirit of the open science framework, we aim to foster increased communication, collaboration, and replication of findings across political science labs utilizing psychophysiological methods.**

Key words: Psychophysiology, Research Tool Report, Skin conductance

The past decade has seen a rapid increase in the number of studies employing psychophysiological methods to explain variation in political attitudes and behavior (e.g., Arceneaux et al., 2018; Bakker et al., 2019; Bakker et al., in press; Carlson et al., 2020; Gruszczynski et al., 2013; Hibbing et al., 2014; Oxley et al., 2008; Renshon et al., 2015; Smith et al., 2011; Soroka et al., 2019; Wagner et al., 2015). This first wave of physiological research was published in high-profile journals (including *Science* and *Behavioral and Brain Sciences*), cited heavily in subsequent work, and covered extensively by the media. Our objective is to maximize the effectiveness of subsequent research in this growing field by demonstrating the utility of psychophysiological methods and providing guidance on how these methods can be employed and consumed most fruitfully.

It is easy to see why research utilizing physiological measures has found an enthusiastic audience. The present theoretical context in political behavior research is well suited to the advantages of physiological methods. An increasing number of scholars highlight the importance of emotion in shaping political behavior, both for ordinary voters trying to understand present political circumstances (Marcus et al., 2000) and for candidates

doi: 10.1017/pls.2020.5

Correspondence: Jaime Settle, College of William & Mary, Government, Tyler Hall, Room 318, 300 James Blair Drive, Williamsburg, VA, 23187; Email: *jsettle@wm.edu*

trying to use emotional appeals to their advantage (Brader, 2006). Closely related research on candidate evaluation emphasizes "hot cognition" (Lodge & Taber, 2013), the notion that all sociopolitical concepts in memory are affectively charged and that these affective responses are brought to mind automatically when a political concept is evoked.

Whether examining discrete emotions or automaticity, one of the core challenges for this research is accurate and meaningful measurement of emotional response. Most traditional research in political psychology relies on the self-reported attitudes of individuals. If we assume that only conscious attitudes have demonstrable effects on political behavior, relying solely on these self-reported data is a sensible approach to tackling questions surrounding individual-level political psychology.

However, the traditional approach of asking people to self-report their emotional states has been called into question by scholars arguing that a great deal of emotional processing happens outside the bounds of conscious awareness (Lodge & Taber, 2013; Smith et al., 2011). These scholars posit that self-reports are not "wrong" but rather are incomplete characterizations of a person's emotional state, as a great deal of work has been done showing that nonconscious attitudes have significant effects on political behavior. That said, emotion researchers do not all agree that psychophysiology can measure discrete emotions (LeDoux 2015a, 2015b), and there is a loose correspondence between psychophysiological reactions, self-reports, and behavior (Lang, 1968; MacDuffie et al., 2019). Like every other field of inquiry, the precise nature of what psychophysiology measures and what it does not is an ongoing area of research. Nonetheless, physiological responses provide indicators of automatic processes that may underlie emotional reactions in politically relevant contexts, and measures that tap into arousal of the sympathetic nervous system can offer a complement to the traditional approaches that political scientists have deployed (Wagner et al., 2015).

There are many ways to measure nonconscious attitudes, including the relatively simple and inexpensive implicit association test (see, e.g., Ksiazkiewicz & Hedrick, 2013 and their edited symposium in *PS: Political Science & Politics*), the electroencephalogram (EEG) test (e.g., Boudreau et al., 2009), and the considerably more complex and expensive functional magnetic resonance imaging technique (fMRI) (e.g., Dawes et al., 2012). Psychophysiological measures such as electrodermal activity (EDA), electromyography (EMG), and pupillometry (eye tracking) provide a midground among these techniques in terms of both complexity and cost (for a discussion of these techniques outside the social sciences, see Lohani et al., 2019). These methods should be an essential part of the political psychologist's toolbox as they provide a cost-effective means of objectively capturing nonconscious, affective responses to environmental stimuli.

While it is clear that interest in physiological methods is growing, several obstacles stand in the way of these tools reaching the full range of scholars who could use them to improve their research. The most obvious of these is a general lack of familiarity with the psychophysiological approach within our discipline. Scholars of political behavior are accustomed to working with many kinds of data, but the technical challenges associated with learning about the sympathetic nervous system and the properties of different physiological measures can be daunting. There are also a number of logistical factors that can discourage researchers, particularly setting up their own physiology labs and buying expensive equipment.[1] Labs must also invest in proper training for research assistants to administer the measures. With physiological research, recruiting participants can also be a challenge, because the methods require researchers to physically interact with participants (e.g., placing electrodes on the skin) and may require higher compensation. Finally, once researchers collect data, they must know how to analyze and interpret them to draw meaningful conclusions.

Beyond these logistical concerns, physiological researchers must also engage with questions of how to place their findings into a larger context. Working in a nascent field of inquiry gives us an opportunity to establish best research practices that are conducive to systematic inquiry. Concerns about transparency in terms of both procedures and theoretical development have arisen across the social sciences, and an increased emphasis on preregistration of studies and replication of findings has followed. In many domains, this sea change has been labeled a "replication crisis" because failure to replicate canonical studies has required fundamental rethinking of what we know. But failure to replicate is

[1]The equipment necessary to measure physiological responses has become much more affordable over the past decade, but it still remains a substantial expense. Costs vary widely depending on the exact configurations purchased, but a base system costs in the ballpark of $10,000. Stimulus delivery software can cost approximately $2,000 more. However, given the popularity of physiological measurement within the field of psychology, many departments in that field may own the equipment, opening the door for fruitful collaborations.

Table 1. Glossary

AcqKnowledge software:	Software program used to collect, measure, and analyze physiological data. This software is similar to packages such as Infiniti from Thought Technology Ltd.
BIOPAC systems:	BIOPAC is a data acquisition hardware used to measure physiology. Other companies include Thought Technology Ltd.
Electrocardiogram (ECG):	An electrocardiogram records the electrical impulses that stimulate the heart. This is done by placing electrodes on a participant's limbs and chest. An ECG can measure heart rate, heart rhythms, and characteristics of the heart chamber (Cacioppo et al., 2007).
Electrodermal activity (EDA):	Electrodermal activity comes from sweat that is controlled by the sympathetic nervous system. When there is a psychological or physiological arousal, there is heightened sweat gland activity. EDA measures levels of skin conductance (Cacioppo et al., 2007).
Electrode:	An electrical conductor that is placed on a participant. An electrode is where electricity would exit the body and into the lead, which then is connected to hardware like BIOPAC. There are many different types of electrodes depending on the type of physiological data that are being collected and the placement on a person's body. For example, with a facial EMG, researchers will want to be cognizant of the size of the electrode and its durability (Cacioppo et al., 2007).
Electroencephalogram (EEG):	An electroencephalogram records electrical activity in the brain. Electrodes are placed on the scalp of the participant. EEGs are known for their high temporal resolution but do not have good spatial resolution.
Electromyography (EMG):	An electromyography measures the electrical movement in the muscles. In the case of a facial EMG, certain muscles in a person's face can denote positive or negative affect. Electrodes are placed on these muscles to measure movement.
Interstimulus interval:	The interval between one stimulus and another. The purpose is to allow participants to normalize or stabilize.
Lead:	The cable that connects the electrode to the hardware.
Nonspecific skin conductance response (NS-SCR):	A type of SCR that occurs in the absence of a known stimulus.
Phasic skin conductance response (SCR):	Phasic SCRs are shifts in SCLs and are usually in response to a stimulus. They are better known as "event-related" SCRs and are thought of as "peaks" in SCR (Cacioppo et al., 2007).
Skin conductance level (SCL):	Tonic level of electrical conductivity of skin (see Tonic SCL).
Stim-Response Analysis:	A type of analysis that can be done through the AcqKnowledge program.
Tonic skin conductance level (SCL):	The skin conductance level of a person without any stimulus or changes in the environment. This SCL does not change quickly.

only a crisis if theoretical claims outpace empirical results. In the area of political psychophysiology, we are not yet at that point. Indeed, recent efforts to replicate early physiological findings (e.g., Bakker et al., in press; Osmundsen et al., 2019; Soroka et al., 2019) have not confirmed the conclusions of those earlier studies. But far from a crisis, these recent efforts have demonstrated the value of the very procedural transparency that this article aims to facilitate and have stimulated healthy discussion over the substantive interpretation of physiological results.

The challenges are real, but they should not discourage researchers from adopting tools that can substantially help their research programs. A better solution is for experienced researchers, who have already endured the start-up costs and begun to wrestle with the thorny theoretical issues described above, to provide guidance to minimize these costs for future scholars. The aim of this article is to provide just this sort of guidance.

A physiological primer

Psychophysiology encompasses a variety of techniques and measures and can be used to describe any study in which the physiological predicts the behavioral, or vice versa (Stern et al., 2001). While the most commonly measured physiological response is electrodermal activity, researchers also frequently collect measures of heart rate, heart rate variability, facial muscle electromyography, and respiratory rate. While less common in political science, others measures include EEG, fMRI, functional near-infrared spectroscopy (fNIRS), pupillometry, and hormonal measurements.[2] Here, we review the physiological underpinnings of these measures. Table 1 provides a glossary of terms that may be useful as a reference in subsequent sections of the article. In the next section, we focus our attention on electrodermal activity and facial muscle electromyography, because these techniques have been put to use most frequently by political scientists studying emotional responses.

[2]While we cannot provide a clear ordering of these techniques based on ease of use or cost, we think that the most approachable measurements for beginners are electrodermal activity and heart rate, followed by hormonal measures and respiratory measures. The most complicated measurements—fMRI and fNIRS—are also among the most expensive.

https://doi.org/10.1017/pls.2020.5 Published online by Cambridge University Press

Key physiological concepts

Electrodermal activity is a measure of electrical conductivity responsive to sweat gland activation, a robust and well-validated measure of physiological arousal. This arousal occurs in response to most forms of novel external stimulation and could be indicative of any affective state: EDA increases when people feel aroused (e.g., anxious, disgusted, startled, angry, happy, proud, interested) in response to a novel stimulus (e.g., a photo). Changes in EDA may also capture changes in attention level (see Soroka et al., 2019). Inferences from shifts in EDA, therefore, are dependent on the nature of the stimuli presented to participants (Dawson et al., 2007). The two principal measurements of EDA are skin conductance level (SCL) and skin conductance response (SCR). SCL represents the overall total of electrical conductivity produced, whereas SCR is a measure of significant deviations from a person's normal SCL.

Electromyography is the label applied to measures of skeletal muscle activation. EMG can measure the activation of any muscle, but psychophysiologists tend to focus on facial EMG, which, as the name implies, is a measure of activation in the facial muscles. EMG is an excellent companion to EDA because activation of various facial muscles acts as a reliable measure of affective valence. By placing electrodes on the appropriate facial muscles, EMG captures either the positive or negative valence in people's emotional responses. For instance, a furrowed brow, a common sign of a negative affective state, involves activation of the corrugator supercilii. Therefore, researchers can infer that study subjects' affective reaction to a stimuli is negative if electrical activity in their corrugator is greater than their baseline activity. Similarly, EMG can capture the degree to which subjects are startled by a stimulus by measuring activity in the orbicularis oculi, which is activated when people blink (blink amplitude increases when people are startled). Disgust has been measured by EMG on the levator labii superioris, a muscle found on the side of the nose between the mouth and eye (Whitton et al., 2014). On the opposite end of the affective spectrum, EMG can also measure the activity of the zygomaticus major, which is associated with smiling and positive affective responses (Tassinary et al., 2007).

Measuring EDA and EMG is an effective way to capture variation in different facets of the nervous system. EDA is part of the autonomic nervous system (ANS). The ANS is responsible for the numerous bodily functions over which people exert little to no conscious control (e.g., sweating, heart rate). The ANS is divided into two major divisions. The sympathetic nervous system (SNS) is the fight-or-flight system, and the parasympathetic nervous system (PNS) is the rest-and-digest system. These systems complement each other in critical ways. Imagine being cut off in heavy traffic: the body detects a threat and prepares itself to respond. Consequently, the SNS activates, causing the heart to beat faster, the palms to sweat, and alertness to increase. However, the body cannot continue in an overly activated state without serious health consequences. Once the threat passes, the PNS brings the body back to its natural resting state. Note that EDA specifically measures activity in the SNS and is not an indicator of PNS activity (Stern et al., 2001).[3] Because people have little control over the SNS, they have little control over EDA responses.[4] As a result, EDA provides more objective measure of how people respond to the world, minimizing biases such as social desirability that bedevil explicit survey measures.

EMG is a measure of the somatic nervous system, which is responsible for voluntary movement and controls the skeletomuscular system. Each of the muscles mentioned earlier—the corrugator, orbicularis, levator, and zygomaticus—are easily controlled by conscious thought, as most people can blink, smile, and furrow their brows at will. Although these muscles are under voluntary control, they nonetheless indicate automatic processes involving little to no conscious control. The orbicularis and blinking are the prime examples, as blinking tends to be a nonconscious activity. When people are startled or disgusted, micromovements in these muscles can be recorded in microseconds with EMG, offering a window into nonconscious processes (see Tassinary et al., 2007).

Physiological hardware and software

Recording physiological data requires equipment, which requires a one-time, up-front expense, albeit a substantial one. Most of the political physiology labs currently in operation require at least two computers: one to present stimuli to participants and another to acquire the physiological data, though newer technology

[3] Some measures of the cardiovascular and respiratory systems, namely, respiratory sinus arrhythmia, capture variation in the PNS (Grossman & Taylor, 2007).

[4] People also have little control over the PNS. We say "little control," because with intensive training, people can exercise *some* control over their ANS (e.g., a trained Navy SEAL), but absent conditioning, people have practically no control over their initial ANS responses.

in physiological measurement is able to operate with only one computer. These computers communicate with specialized equipment—proprietary hardware and software licensed and distributed by a number of companies, such as BIOPAC or Thought Technologies—to acquire and process physiological data. Researchers place electrodes on participants that pass small levels of electrical current to assess the raw electrical signals coming from participants. These electrodes are attached to amplifiers (via wires or wireless connections), which, in turn, connect to the data acquisition computer (via an ethernet cable), which then translates the raw electrical signal produced by participants into data captured with a specialized software program, such as BIOPAC's AcqKnowledge program or Thought Technology's Infinit platform. Stimulus delivery software programs, such as SuperLab or ePrime, can be configured to communicate with physiological software programs, permitting exceptionally precise indicators of stimulus onset.

Best practices in psychophysiological research

Political psychophysiological studies share many things in common with other forms of laboratory studies: the need for transparent research designs, the importance of clear documentation for protocol procedures, and the challenge of balancing experimental control while striving for generalizable findings. However, the nature of psychophysiological research exacerbates the importance of some features of laboratory experiments—such as the need to control the physical conditions of the lab—and introduces challenges not typically associated with laboratory experiments—such as the variety of post-processing decisions on voluminous data sets. In this section, we provide an overview of the considerations that are magnified in importance when conducting psychophysiological research.

Here, we draw on four principles—transparency, systematic methodology, replicability, and robustness—to derive best practices. Because selecting the "right" approaches to cleaning psychophysiological data and constructing measures is often question-specific and a subject for ongoing methodological research (e.g., Boucsein, 2012), we do not wish to take positions on the "right" way to inspect data or derive measures. Instead, we offer some guidelines about what scholars should consider when collecting and analyzing physiological data as well as evaluating research that employs these measures.

Transparency requires researchers to clearly describe the way in which they record, recode, and analyze their physiological data. Because there are many research degrees of freedom (e.g., there are dozens of ways researchers could identify outliers), we encourage researchers to preregister their protocol and data analysis plans. *Systematic methodology* requires that researchers justify the decisions that they make when it comes to the collection and analysis of physiological data (ideally in a preanalysis plan) and apply those decisions in a systematic way. *Replicability* requires that other researchers could follow the steps that researchers describe for collecting and analyzing data to produce the same analytical outcomes (e.g., identify the same outliers, produce the same measures). Finally, *robustness* requires that the conclusions that researchers draw from the analysis of physiological data are consistent across alternative analytic decisions (e.g., a different but reasonable method for identifying outliers or constructing a measure from raw data).

Research design

As with all laboratory studies, many of the most important decisions are made at the design stage. In addition to all of the principles guiding strong research design generally, research in this area is also guided by the possibilities and limitations of physiological measurement. In this section, we highlight some of the areas in which these two sets of guidelines reinforce one another or create tensions.

Experimental design. Most of the early works in political psychophysiology are observational studies rather than experiments. The stable, biological basis of physiological dispositions limits experimental control in many instances. We cannot, for instance, randomly assign threat or disgust sensitivity to study participants. Therefore, in these early studies, researchers collected physiological measurements to study correlations between political attitudes and physiological measures of interest without making strong causal claims about the effects of physiological response.

Among the reasons for the continued importance of studies without clear control groups is the issue of statistical power. Physiological studies require intensive human capital, making it difficult to obtain enough subjects for experiments with multiple treatments. Furthermore, in our experience, data loss rates can be as high as 20% for physiological measurements such as EDA

https://doi.org/10.1017/pls.2020.5 Published online by Cambridge University Press

and EMG.[5] Traditionally, within-subjects designs have been used to get around this issue because they allow researchers to include multiple trials in one study (e.g., one for inducing anxiety, another for anger). In fact, in physiological research, within-subjects designs may be preferable for several reasons. External factors such as temperature and humidity can influence subjects' physiological responses. For this reason, within-subjects analyses are often the most appropriate, as all changes can be attributed to the subject's physiology rather than differences in conditions in the lab. Within-subjects analyses also have the added benefit of accounting for the large differences in physiological responses between subjects (e.g., some people have thicker skin, dampening EDA response), which can make it difficult to balance physiological reactivity across treatment groups in small samples.

However, there may be research questions when within-subjects designs are not practicable or researchers have strong hypotheses about how different treatments should evoke different physiological responses (e.g., Mutz & Reeves, 2005; Soroka, 2014). In these instances, between-subjects designs are feasible, but they are best when researchers are able to conduct meaningful power analyses in advance. Effect sizes are often small, making it difficult to identify statistically significant results, and, as with all power analyses, researchers should be careful to look to previous research using similar designs to estimate realistic effect sizes. Ideally, these designs should always contain some common stimuli between groups to establish baseline physiological reactivity equivalence between groups, and researchers should also collect data on participants' baseline physiological responsiveness to a stimulus with known physiological response properties.

Additionally, as with all studies using stimuli, pretesting is essential. Whenever practicable, and always when using a novel stimulus or a novel protocol, it is important to assess subjects' perceptions of the characteristics and interpretation of the stimulus before beginning the costly process of measuring physiological response. For example, Smith et al., (2011) pretested a variety of images in order to differentiate between neutral stimuli, disgusting stimuli, and other negative (but not disgusting) stimuli. This careful image selection allowed them to separate the impact of disgust from a general negative emotional valence.

Stimuli and measurement selection. The preponderance of studies in the field so far have used nonpolitical but well-validated stimuli to evoke physiological responses. The pioneering work (Oxley et al., 2008) of the University of Nebraska–Lincoln Political Physiology Lab, for example, collected physiological measures on well-established protocols such as the International Affective Picture System (IAPS). Using well-validated stimuli (e.g., Bradley and Lang, 1999; Bradley and Lang, 2007; Lang, Bradley, and Cuthbert, 2008) known to evoke particular physiological responses associated with particular emotional or cognitive states is an important part of establishing a role for physiological response in political behavior. These studies suggest a link between a person's generalized physiological response and his or her political attitudes or behavior but do not necessarily test physiological response *to* the political environment per se.

Other work has extended into testing physiological response to explicitly political stimuli. This approach represents a conceptual shift: instead of arguing that generalized physiological sensitivity is associated with particular political attitudes or behaviors, this research implicitly asserts that physiological response to the political environment itself is an important and meaningful behavior to measure. One of the earliest contributions was that of Mutz and Reeves (2005), who measured EDA response to watching civil and uncivil depictions of candidate debates, finding that watching uncivil interactions elevated EDA. The advantage of using political stimuli is the ability to capture an implicit measure to complement what is already known about explicit self-reported responses to the same or similar stimuli. However, without differentiating whether subjects are sensitive to particular classes of stimuli more generally (e.g., anger-inducing stimuli) it is difficult to separate generalized sensitivity from sensitivity to a political stimulus. Researchers should adapt their designs accordingly if they want to make such a claim.

In addition, while it is always important to consider the order in which stimuli and tasks are presented to a study participant, within physiological studies, the considerations extend beyond concerns of priming and order effects. It is essential to consider what effect each stage of a protocol will have on the participant's subsequent mental states or physiological responses. For instance,

[5] In our experience, we have found data loss rates of up to 15% for EDA (Carlson et al., 2019) and 10% for other measurements, such as EMG. Other researchers, however, such as Bakker et al., (2019), find greater data loss rates for EMG than EDA. Data loss rates are likely to vary from one study to the next, but the broader point is that researchers should be prepared to lose some data when collecting physiological measurements, and this should be a consideration in power analyses.

hooking a participant up to the physiological equipment may change her or his emotional state; the interaction between a proctor and a subject is much more intimate than in a standard study, which could be a relatively stressful experience for some subjects. One good (but time-consuming) method of dealing with this potential anxiety is to allow for an acclimation period before measuring baseline states or presenting stimuli. During this acclimation period, researchers can ask subjects to complete innocuous tasks, such as reading protocol instructions or answering basic demographic questions. Assuming a calm environment, it should only take a few minutes for participants to return to a relaxed state, thus allowing the researcher to establish a baseline state.

Many physiology labs collect multiple different measurements during a single protocol in order to test several hypotheses at one time and to use multiple physiological measures in conjunction to better characterize subjects' response. However, as elaborated later, physiological recordings can be compromised by interference between physiological measurements (Cacioppo et al., 2007; Stern et al., 2000). For instance, EMG signals can introduce artifacts into EDA signals. A helpful chart of how different psychophysiological indicators interact to produce such artifacts can be found in Cutmore and James (1999, p. 135).

The imperative for laboratory control. Regardless of the stimuli used or the kind of physiological data collected, it is important to standardize the participants' experience and to maximize participants' comfort and focus. Any variation in protocol from participant to participant may cause changes in participant response to stimuli. Additionally, if the laboratory setting causes a participant undue stress, that stress will manifest physiologically, making it impossible to separate responses to the setting from responses to the stimuli.

Researchers can reduce artifacts when measuring EDA by maintaining proper lab conditions, adhering to a strict experimental protocol, and providing clear instructions to subjects. Ideally, stimuli should be presented and experimental tasks should be performed in otherwise quiet (sound-attenuated, if possible) rooms, with few visual distractions and low traffic. It is important that the lab be climate controlled to maintain relatively constant temperature and humidity, especially when measuring EDA (Boucsein et al., 2012), because EDA measures tend to be positively correlated with ambient temperature (Boucsein et al., 2012). If possible, labs should maintain a room temperature of around 22–24 degrees Celsius, or 72–75 degrees Fahrenheit (Braithwaite et al., 2013, p. 41). It may be useful to present auditory stimuli via headphones to reduce distractions from other noises.

It is essential that researchers codify all decisions about lab set up in a protocol. Whenever possible, proctors should adhere to a script when relaying instructions to participants to ensure consistency between subjects and to standardize the lab experience. The more the protocol addresses, the better. It is advisable to include full instructions in the protocol addressing details from before the time of participant arrival—including setting up the lab and turning on equipment—until after all participants have left and the lab is to be shut down. Such a protocol facilitates replicability and should be used when training new proctors. Because of the sensitivity of physiological measurement, seemingly trivial differences in the way proctors administer the experiment (such as the placement of electrodes) can have downstream consequences for the interpretation of results.

Study length and participant fatigue. Like all other experiments, participant fatigue is a concern when measuring physiological responses. Researchers should be mindful of the length of their studies and may want to schedule short breaks into the experimental protocol if feasible. There is no established maximum time duration for physiological data collection, as study length necessarily depends on the hypotheses, stimuli, and design unique to each study. However, based on our experience, we suggest that researchers aim for no more than 30 minutes of data collection to avoid stress or fatigue. While this is not a hard-and-fast rule, study designs lasting considerably longer than this recommendation should be justified in light of duration-related concerns posed in these types of studies. If a researcher wants to collect data for multiple hypotheses at one time, experiments can become quite lengthy. As we discuss below, long protocols may introduce artifacts into the data.

First, the likelihood that a participant will move during a study increases over time. As explained later, movement can introduce noise into the data. Second, if participants are exposed to repetitive stimuli, habituation can occur (Dawson et al., 2001). This habituation will manifest as a general decline in amplitude of response over time. In other words, while novel stimuli can still evoke arousal, the longer participants engage in the same task, the less responsive their SCRs will be. To our knowledge, there is no set length of time at which habituation will occur or become a limiting obstacle in

https://doi.org/10.1017/pls.2020.5 Published online by Cambridge University Press

research designs. Habituation itself is influenced by individual differences; for example, it is highly correlated with psychological resiliency and other individual differences (see Walker et al., 2019).

The type of stimuli used and the degree to which participants begin to detect and anticipate patterns in their presentation will also influence habituation. Relatively neutral stimuli such as images may lead to faster rates of habituation than startling stimuli, such as loud pulses of sound. Just as one would with any study in which respondents are exposed to multiple stimuli,[6] we suggest that researchers account for habituation in the research design stage by carefully considering stimuli design and potentially randomizing the order of stimuli that are presented to subjects.

Another concern when measuring EDA during longer experiments is *drift* in the background SCL over the course of a study. Drift is a general increase in the background tonic SCL (Braithwaite et al., 2013, p. 11) and can occur for a number of reasons, including unreliable connections between electrode attachments, data recordings that produce time lags, disproportionate use of electrode gel, electrode polarization, or even a buildup of sweat over time (Shaffer et al., 2016). It is sometimes difficult to ascertain whether changes in SCL are due to drift, habituation, or substantively meaningful physiological responses. To account for these artifacts, we recommend implementing baseline measurements in advance of each stimulus, what are called interstimulus intervals. These are typically designed as blank or black screens, potentially with a cross or short bit of static text to focus participants' attention, of 30-60 seconds. Because meaningful physiological responses only take seconds to manifest, while drift or habituation may take tens of minutes (Braithwaite et al., 2013, p. 11), using these rest intervals to establish stimulus-specific baselines should account for any artifacts in SCL measurement.

Ethics review. Institutional review boards (IRBs) vary widely from institution to institution, and psychophysiological laboratories at different universities have had radically different experiences with feedback from the IRB. Some IRBs have been particularly concerned about the nature of the stimuli that are being shown to subjects (such as IAPS images designed to elicit disgust), while other IRBs have been more concerned about the invasiveness of the procedures to attach electrodes or other measurement equipment. Our general recommendation is to factor in additional time for ethics review, to allow researchers to address any concerns an IRB might raise. We also encourage researchers to treat responsibly the fact that they are collecting measurements that can be informative about subjects' health and well-being. They should be clear on an institution's policy for the disclosure of information to a subject about a measurement that indicates a potential health issue.

Data collection

Data collection presents challenges unique to psychophysiological research. In this section, we discuss these challenges and explore possible solutions. In a laboratory study, researchers may collect EDA, EMG, and other physiological responses, such as heart rate and respiratory readings. Each requires different hardware with different attendant concerns. Given space constraints, we focus primarily on EDA in this section because it has been the measure of psychophysiological response most frequently employed by political scientists, but we make brief reference to other types of measurement as well.

Advance preparation. Unlike many other types of lab studies, researchers need to collect information from subjects in advance of the study itself, making a presurvey essential. Because physiological equipment is able to detect changes in physiological response over a millisecond, it is important to properly screen participants for any circumstances that may affect measurement. Experimenters should send a reminder to subjects before their scheduled appointment to discourage behaviors that may interfere with the measurements (e.g., use of tobacco, alcohol, or hand lotion).[7] Participants should also be reminded to wear clothing appropriate for the experiment.[8]

Electrodermal activity configuration. High-quality EDA data can be difficult to record. Readings rely on a good electrical connection to an unmoving, focused

[6]For example, consider a study that assesses self-report emotional evaluations of multiple campaign advertisements. If researchers are interested in testing whether subjects have different emotional responses to different advertisements, they would randomize the order of the advertisements to account for ordering and habituation effects.

[7]Hand lotion is problematic because it alters the hydration of the skin, which alters its conductivity for the purposes of EDA measurement. Tobacco, caffeine, and alcohol can affect heart rate. In situations in which it is not feasible to discourage these behaviors (see, e.g., Bakker et al., 2019), researchers should justify the decision in a preanalysis plan.

[8]Heavy sweatshirts or sweaters may impede the collection of respiratory data, and leggings or stockings may create awkward situations when connecting ECG leads to subjects' ankles.

https://doi.org/10.1017/pls.2020.5 Published online by Cambridge University Press

participant. To ensure the best-quality data, it is essential to prepare the participant adequately before attaching the sensor. We recommend that participants rinse their hands in lukewarm water (and dry them) without soap, as soap may cause swelling of the epidermis and depress tonic EDA readings (Boucsein, 2012, p. 109).[9] Researchers should place the electrodes on the skin at least 10 minutes before baseline measures are established. This allows adequate time for the electrolyte gel to penetrate the outer layer of the skin, ensuring a stronger signal (Boucsein et al., 2012, p. 1023). If subjects wash their hands during a break in the study, the experimenter must reapply the electrolyte gel and allow another 10 minutes for the sweat glands to reactivate before continuing to record the EDA signal.[10]

If the participant moves during the study, or if the electrical connection is disturbed in any way, the data become noisy, making it difficult to establish a link between stimulus delivery and electrodermal response. Because motion can create artifacts when measuring physiological data, subjects should be instructed to keep movement to a minimum. Similarly, coughs, sneezes, or loud, unexpected sounds (such as construction noises or a cell phone ring) can also create artifacts in the data. Many of these are unexpected and thus, unavoidable. However, data acquisition programs allow experimenters to tag such events using hotkeys so that they can be accounted for in the final analysis (Braithwaite et al., 2013, p. 41).

EDA measurement also requires attention to electrode placement. EDA electrodes must be placed in pairs, to create an electrical circuit. These pairs should be placed somewhere on the hands. The exact placement of EDA electrodes varies from lab to lab, as it is largely dictated by particular hardware models and the nature of the experimental tasks. While sites other than the hands are possible (van Dooren & Janssen, 2012), many are invasive (inner thigh, buttocks), awkward (forehead, armpit), or underresponsive (shoulders, calves). Ideally, the electrodes should be connected to the distal phalanges of fingers (Boucsein et al., 2012). However, if an experiment requires participants to type or otherwise use their hands, a fingertip site may be avoided, as jostling the electrodes is likely to result in unusable data. In this case, sites on the medial phalanx of the fingers or on the thenar and hypothenar sites on the palm are preferable. Ideally, the electrodes should be attached to the participant's nondominant hand, as there should be less callous tissue (Boucsein, 2012). If both hands are needed for tasks, consider placing electrodes on the inner aspect of the foot (Boucsein et al., 2012; Edelberg, 1967). If participants are to perform tasks that require both walking and using their hands, the shoulder is a suitable location for the placement of electrodes (van Dooren & Janssen, 2012).

Data collection. At the start of a lab session, care should be taken to ensure robust measurement. Poorly attached electrodes or other technical malfunctions may negatively impact measurement quality. We recommend a brief calibration period before presenting stimuli. This allows for verification that the data for each specific psychophysiological indicator appear normal. If readings appear to be compromised, researchers can troubleshoot prior to data collection. This may be as simple as verifying that adhesive electrodes are firmly attached and all connections are secure. In the case of EDA, poor readings may be the result of inadequate amounts of electrolyte gel being applied to the electrodes before attachment. If one-time disposable electrodes are being used, replacement of electrodes with a fresh set may resolve technical issues.

It is worth noting that approximately 10% of individuals are estimated to be nonresponders (known as hyporesponsive) in terms of their EDA. While we suggest checking all electrode contacts and restarting the data collection software to verify that the problem is not with the signal, it is simply impossible to obtain high-quality EDA measurements from some individuals (Braithwaite et al., 2013, p. 41).

After stimulus presentation has begun, proctors should continue to monitor data collection to ensure that physiological measurements remain robust over the course of the lab session. For instance, adhesive sensors may come loose partway through a session, compromising data quality. If short breaks are incorporated into the research design, this not only combats participant fatigue over the course of a session but also allows for additional troubleshooting to limit the time that poor-quality data are collected. When saving recordings of participants' physiological data, it is helpful to include a short note of any issues on specific physiological indicators that can be referenced during data cleaning or analysis. It is also essential to ensure data are regularly

[9]The swollen corneum creates pressure that can inhibit the pores from opening, which inhibits sweat from reaching the surface of the skin and prevents accurate measurement (Boucsein et al., 2012, p. 1023).

[10]It is not necessary to use an electrode-preparation product to abrade the attachment site. In fact, abrasion is specifically discouraged for exosomatic EDA recording (Cacioppo et al., 2007, p. 163).

backed up in secure locations over the course of a study. Minimizing data loss is crucial given how time and labor intensive the data collection process can be for psychophysiological research.

Given the complexity of the data collection process, we recommend maintaining a log recording pertinent details of a participant's lab session that may affect the strength of the physiological signal. No detail is too small! For example, it is often prudent to ask participants whether they need to use the restroom prior to beginning the lab session, depending on its expected duration. However, if participants use the restroom immediately prior to the start of a lab session, the use of soap afterward to wash their hands will impact the levels of sweat on their skin and therefore their likely EDA measurements. When physiological responses to non-study-related distractors occur, these should be noted both in the software recording the data and in the lab log. Other aspects of a participant's lab session should also be noted, such as excessive fidgeting, drowsiness, or extreme emotional states.

Data analysis

One of the challenges of psychophysiological research is the complexity of the post-processing of the data. While these are often small-*N* studies due to the intensity of the data collection procedures, researchers collect millions of observations for each subject's raw data: measurements for one or more physiological variables with up to a thousand observations per second over the duration of a study, for example. As a result, the decisions a researcher makes in meaningfully aggregating the data are consequential for the interpretation of findings.

Data exportation and sampling rates. To date, most political scientists collecting these data have used the AcqKnowledge software; therefore, we make specific reference to some of its features. However, other software and hardware configurations work similarly. There are two options to extract data from the AcqKnowledge software. The first capitalizes on AcqKnowledge's built-in features. The "Stim-Response Analysis" is an appealing part of the program that works especially well in research designs in which stimuli are given to participants at specific times. These stimuli events are translated to digital inputs that can be transferred into "stim events functions" that are then interpreted as a binary number. This is also helpful when the design includes multiple levels of stimuli events and the researcher wants to analyze the data for the experimental conditions separately. The "Stim-Response Analysis" allows the experimenter to tell the program how many seconds of data to analyze around the stimulus event. Then, the program can extract the type of data necessary for analysis (e.g., mean, min, max, standard deviation). These data are then transferred to a data file, where the information, including the onset times, are neatly printed out for further analysis. The concern with this approach is that it is at odds with the principles of the open science framework: these analyses will not be reproducible for other researchers because the algorithms used by proprietary software are not entirely transparent and are specific to each package's users.

The second way to analyze the data involves exporting it: the data in AcqKnowledge (as well as other software packages) can be saved as a text file that can then be uploaded to a program of choice (e.g., Stata, R). The key benefit of this approach is its potential for consistency with the best practice principles we outlined: transparency, systematic methodology, replicability, and robustness. Additionally, this option allows the experimenter to decide how fine-grained the analysis should be while keeping in mind what the design warrants and how to optimize the data analysis process. Some stimuli may be long—for example, having participants look at a photo for 20 seconds. Some stimuli may be quick and startling. For the first example, in which participants' physiological responses are collected over a long period of time, it may be necessary to extract only 50 data points per second. In the second scenario, in which the immediate reaction is the target response, it may be beneficial to get 1,000 data points per second. Data sampling procedures will, as always, depend on the nature of the study and the specific physiological measures being collected. EDA measures respond slowly, so there is limited benefit to high sampling rates, but other psychophysiological measures may be more sensitive to this choice. Given the numerous combinations of physiological measures and research designs, it is best to refer to the literature surrounding those measures or designs to determine the appropriate sampling rate.

Measuring EDA. Regardless of how data exportation is done, there are a number of factors to consider when creating variables from physiology data. EDA is usually characterized in one of two ways—tonic skin conductance level[11] and phasic skin conductance response. SCL captures participants' current level of arousal and SCR

[11]SCL is usually measured in microsiemens (μS).

captures changes in SCL in response to some stimuli (e.g., a photo or noise). Phasic skin conductance response recorded over a longer interval and not in response to particular stimuli is called nonspecific skin conductance response.[12] Researchers capture SCR through multiple approaches. In one approach, researchers measure participants' baseline SCL during an interstimulus interval, such as a blank screen, that precedes exposure to the stimulus and then measure their SCL during exposure to the stimulus (e.g., a photo). Another approach collects baseline SCL during a 2- to 15-minute period prior to the experiment and then collects SCL during the experiment. Using both protocols, SCR is simply the difference between the baseline SCL measure and the SCL measured during exposure to the stimulus (or stimuli). Getting an accurate baseline is particularly important when there are subgroups of the population that have tonic hyper arousal—high levels of SCL (Dawson & Nuechterlein, 1984).[13]

One important caveat is that some people simply do not exhibit any phasic changes in their EDA. Some researchers have suggested that this is around 25% of the normal population (Venables & Mitchell, 1996).[14] There are also situations in which a person may be responsive on one hand/wrist but not the other. If the nondominant side is unresponsive, it may be warranted to try the other side.

Variable transformations, data cleaning, and data analysis. Visualizing physiology data can do wonders for helping the experimenter understand the data on an individual and aggregate level. For example, habituation can occur when participants are being recorded for long periods of time (Andreassi, 2007), and the best way to detect this is a visual inspection of the data. However, if the research design has evenly spaced intense or complex stimuli, this habituation may not occur. Plotting the physiological outcomes over the course of the experiment should reveal information about the effect the stimuli had on the participant. Looking at these data for an individual participant can be helpful in identifying any outliers or noise and can be compared with the protocol log notes for any irregularities during data collection for that participant. On an aggregate level, a plot can expose the overall trajectory of physiological responses over the course of the experiment. Data visualization can also show whether there are any irregularities in the distribution of the data.

Visualization is an opportunity to easily check for any data quality issues. For example, any negative readings in the raw data most likely indicate a calibration error with the equipment, which may or may not render the data usable. As mentioned previously, there is a substantial portion of the population that is deemed to be "nonresponders." In this case, it is appropriate to exclude them from the data, as there will be no variation in their response. Lastly, this is where diligent and meticulous lab notes are important. It is acceptable to exclude data if there is reason to believe that the readings are unreliable.

If there were abnormalities during the data collection process, it is up to the researcher's discretion to exclude data. Unfortunately, this is an area of physiological research where there is little consensus or firm guidance. Our advice is for researchers to preregister the data cleaning and transformation plan as part of a preanalysis registry in order to develop protocols for excluding participants' data *prior* to data analysis and preferably before data collection. It is key to be fully transparent when reporting results about the reasons participants were excluded. What counts as unacceptable data, how is it recognized, and who makes that decision? A stronger approach could also include a blinding procedure. For example, a research assistant blind to the study design inspects the data to decide which data fall below quality standards according to predetermined metrics or standards. In addition to data quality issues stemming from equipment malfunction, participant nonresponse, or participant noncompliance, Cacioppo et al., (2007) suggest identifying any outliers, as they can skew the data.[15] The research assistant then sends the data to the authors who use the data to test their preregistered hypotheses.

The next step is to transform the data. EDA data often require transformation if an experimenter finds the distribution of data to be skewed or behaving abnormally. Physiological data can be susceptible to skewing, kurtosis, and heterogeneity of variance (Dawson et al., 2001).

[12]SCR is used to describe phasic responses, and NS-SCR refers to phasic increases that occur but are considered tonic because they are measured in the absence of stimuli (Dawson et al., 2001).

[13]Hyperactivity has been shown to be prevalent in people with severe illnesses (Katsanis & Iacono, 1994) and low brain metabolism (Hazlett et al., 1993).

[14]A consistent finding in psychophysiology is that schizophrenia patients are likely to be electrodermal nonresponders. Bernstein et al., (1982) found 50% of schizophrenia patients to be nonresponders.

[15]Outliers are not always problematic. In fact, they can be the measurement of interest particularly when looking at phasic responses. The experimenter must simply ensure that the outlier is not an outlier through error.

There are many ways to transform and clean data. Examples include a natural logarithmic transformation (see Boucsein, 2012; Dawson et al., 2001), taking the square root or standardizing SCL scores (see Boucsein, 2012; Smith et al., 2011), time-series smoothing (Soroka, 2019), and area under the curve (Boucsein, 2012). See Soroka (2019) for the elaboration of an empirical example. We encourage researchers to inspect the transformed data to verify that the distribution is well-behaved (i.e., single-peaked) as well as to conduct sensitivity analyses to ensure that their statistical results do not depend on a particular transformation.

Reporting results

As with the reporting of any laboratory experiment results, the write-up of psychophysiological studies should facilitate critical evaluations of the quality of the research design, data collection, analysis, and the interpretation of results. The amount of detail needed to inform future work in the form of replications and extensions, however, is considerably greater in this area of study, given the sensitivity of measurement and the many decisions involved in the post-processing of data (Fowles et al., 1981). In the following sections, we outline the content that we believe should be included to satisfy expectations for replicability and transparency (JETS, 2015).

Main text. In many ways, reporting the details of psychophysiological studies is similar to that of traditional lab experiments. Here, the report generated by the Standards Committee of the Experimental Research Section of the American Political Science Association provides a good starting point. That document provides a list of elements that researchers should include when conducting laboratory experiments. These elements include eligibility, recruitment, and incentives for participation, characteristics of the sample and relevant covariates, and an explanation of the protocol, including the treatment stimuli and timing intervals used (Gerber et al., 2014, pp. 94–98).

However, protocols in psychophysiological studies in political science are unusually long, detailed, and complex compared with protocols for other lab studies. As such, it can be difficult to discern which details to include in the body of the article and which details to include in the appendices. Beyond standard best practices for reporting data analysis, political psychophysiological studies should consider a few additional points. First, any information that pertains to the generalizability of the evoked physiological response—such as additional details regarding the demographics of the sample, information about the delivery of the stimulus, or lab conditions—should be included. Also, there is a relatively high expected rate of data loss in this field of study. There are many reasons why participants might need to be removed from a psychophysiological analysis: equipment malfunctions, proctor error, or extreme outliers in terms of physiological reactivity or hyposensitivity. Researchers should clearly report what proportion of participants were removed from the analyses, explain and justify the criteria used to remove participants from the analyses, and report the sensitivity of results to these exclusion criteria.

As discussed earlier, many decisions are required to operationalize psychophysiological concepts into data suitable for analysis. When reporting results, it is critical to clearly explain the substantive interpretation of the physiological measures, how data were aggregated, what transformations were applied, and how summary statistics were calculated. The transformation of physiological variables often makes it somewhat difficult to conceptualize the substantive impacts of stimuli on physiological response. We recommend that whenever possible, researchers visualize their results and include simulated effects of first differences or ranges of the physiological measure to aid interpretation. Researchers can also provide information about physiological response in populations similar to their samples or to stimuli similar to their own to aid readers in contextualizing results.

Appendices. Unlike other studies, we recommend that researchers make available their full, detailed lab protocols in supplemental appendices rather in the main text. These protocols should include detailed information about what happened when participants arrived at the lab, how they provided informed consent, when and how they interacted with proctors (including specific scripts used by proctors to interact with participants), when and how the psychophysiological equipment was applied to the participants, the timing of stimuli, how participants were debriefed, and how data were processed after collection. Ideally, the full treatments should be available in the appendix if feasible. If not, the appendices should at least include sample graphics of the treatment to help readers visualize exactly what participants saw or read in the study. If audio or video treatments were used, researchers should make every effort possible to make them available in an online appendix. Finally, researchers should also make proctor-training materials available upon request. For an example, albeit an imperfect one, see the online appendix of Carlson, McClean, and Settle (2020).

Because of the high setup costs involved, it is common practice to collect data for multiple studies in a single data collection session. We believe it is important to provide at least summary information in the appendices about the full set of physiological recording data measured on a subject, even those that are not utilized in the analysis. Ideally, researchers would also preregister analysis plans that explicate their hypotheses, exclusion criteria, measurement decisions (e.g., transformations), and statistical modeling decisions (Miguel et al., 2014). These steps would help address the concerns raised by Franco et al., (2014) about publication bias.

Null results. Although publication bias against null results exists in the social sciences (Franco et al., 2014), we encourage researchers to submit their null results, in the main text or the appendix as appropriate. Similarly, we encourage reviewers to be open to null results when critically evaluating work for publication. As psychophysiological studies are particularly time-consuming and expensive to conduct, the imperative to publish null results from well-designed and well-executed studies becomes even more important to prevent wasted time and resources. Researchers can examine null results that they find interesting and build on them, make research design improvements, and otherwise seek to extend and replicate published works to push the field forward. It is for this reason that we believe it is important for labs working on similar questions to communicate results with one another. Preregistration of analysis plans and research reports offer a useful mechanism for doing so.

Caveats about psychophysiological measurement

Clearly, we see a number of benefits to be gained from incorporating measures of psychophysiology into political science research. Nonetheless, we appreciate the limitations that are inherent in the approach. While these issues do not dampen our optimism, we think it is important to present a complete picture. No measurement approach is perfect and psychophysiological measurement is no different. We discuss two broad caveats that scholars should consider before incorporating psychophysiological measures into their own research as well as when consuming research produced by others.

Sampling limitations

At this time, psychophysiological measurement requires researchers to have direct interaction with study participants under controlled conditions. The best way to accomplish this is for researchers to bring their participants into the laboratory. Although the technology of measuring physiological responses has become more available and more affordable, the equipment remains a considerable expense, especially when compared with standard social scientific measurement tools. Consequently, most social science laboratories can generally only measure the physiological responses of a handful of respondents (generally, one or two) at a time.

These practical limitations translate into sampling limitations. Most psychophysiological studies rely on small samples, often drawn from conveniently located populations such as college students. These sampling limitations raise two important concerns. The first is that underpowered studies, when combined with standard frequentist approaches to null hypothesis testing, are more likely to exaggerate true effect sizes, since small samples can only "detect" (i.e., generate small p-values) large effects. As we noted earlier, the findings from several early and influential political psychophysiological studies have failed to replicate when repeated with larger samples (Bakker et al., in press; Osmundsen et al., 2019; Soroka et al., 2019), highlighting the need for larger samples. The second concern is that even if one isolates the true relationships between physiological measures and outcomes of interest, they may only apply to a narrow population and therefore fail to generate knowledge generalizable to the broader populations in which political scientists tend to be interested.

It is important to note that while these concerns often accompany studies that incorporate psychophysiological measurement, they apply to all social science research. Indeed, laboratory research and convenience samples are quite common in other disciplines such as neuroscience, psychology, and behavioral economics. Perhaps more importantly, research into the validity of lab experiments finds that results from such studies have a strong correlation with results from both online (Clifford & Jerit, 2014) and field studies (Coppock & Green, 2015). Furthermore, comparison studies find that student samples perform similarly to probability samples, especially when the treatment effect does not depend on a characteristic for which there is little variance (Coppock et al., 2018; Druckman & Kam, 2011). Regarding psychological traits and political attitudes specifically, research regularly

https://doi.org/10.1017/pls.2020.5 Published online by Cambridge University Press

finds few, if any, systematic differences between student and representative samples (Vitriol et al., 2019).

Still, given the persistent concern over the validity of convenience samples, we believe that the best response to this limitation in psychophysiological research is the same that scientists use elsewhere: replication. No single study, not even ones based on large and representative samples, can be taken as definitive. In fact, the entire edifice on which null hypothesis testing is built presumes that we only gain knowledge about true effect sizes through multiple studies. We believe these considerations offer another rationale for psychophysiology labs to work in concert. A network of labs can facilitate both larger-*N* studies and replication (Klein et al., 2014).

We must also keep in mind that technology is not static. As more social scientists incorporate psychophysiological measures into their toolkit, we will develop more efficient and effective protocols. Moreover, as anyone with a recent vintage smart watch can attest, the technology for measuring physiological states (e.g., heart rate) can be miniaturized and made more available. Consequently, there may be a time when researchers do not need a lab filled with expensive equipment to record at least some psychophysiological measures.

Substantive interpretations

Our second major caveat applies to the substantive interpretation of psychophysiological measures. As we noted at the outset, EDA measures arousal, and researchers often use it as a marker for emotional response. However, EDA cannot be directly mapped to specific emotional states, and some researchers also consider EDA to measure attention (e.g., Soroka, 2019). Disparate emotional states, including anxiety, anger, and excitement, can generate increased dermal activity. Consequently, psychophysiological measures cannot and should not be used in the same way that social scientists often use off-the-shelf self-reports, which are designed to measure specific psychological traits and states. As is the case with any measurement approach, researchers must take care to ensure minimal slippage between the background concept they wish to measure (e.g., anxiety) and the operationalization of that measure (e.g., self-reports, EDA). With respect to psychophysiological measures, researchers must interpret physiological responses in the context of their research design.

Moreover, the same issue presents itself with self-reports. For instance, a recent study finds that increases in EDA in response to disgusting images does a better job predicting political attitudes than does people's self-reported levels of disgust with the same images (see Smith et al., 2011). People do not always have access to their emotional states—or worse, they may wish to hide them from researchers—rendering self-reports imperfect measures as well. The point here is not that psychophysiological measures are always better or preferable to self-reports. It is that they offer social scientists additional leverage at measuring important psychological concepts. We view them as an addition to our methodological toolkit, rather than a replacement of existing tools.

Conclusions

Collecting psychophysiologically informative data can be costly and time-consuming, but the potential value added from these methodological approaches can complement and extend research on a broad set of questions of interest to political scientists. We are at a critical juncture in this area of study, as the community of interested researchers has the opportunity to come together and develop rigorous standards for transparent, systematic, replicable, and robust research. Given the immense potential of psychophysiological research to contribute to the discipline, we hope that it will not become siloed. Our goal in compiling this primer is to lower the barriers to entry for researchers interested in (1) incoporating these techniques into their research or (2) serving as informed, constructively critical reviewers of research using these approaches.

We summarize the main points from our extended discussion as follows:

- Researchers should rely on theoretical guidance before measuring psychophysiology. For example, are researchers interested in measuring physiological arousal, emotional states, coping ability, or something else? Answering these questions first will guide researchers to appropriate physiological measures. We provided some reference citations at the beginning of this article that will help researchers navigate these questions.
- Researchers should develop and follow standard protocols for administering physiological testing. These standards include settings for the biosensory equipment, protocols for handwashing, connecting biosensors, and helping participants relax before the beginning of the session and remain still during the session.

https://doi.org/10.1017/pls.2020.5 Published online by Cambridge University Press

- Researchers should develop transparent criteria for data post-processing and analysis. This includes procedures for inspecting raw physiological outputs for anomalies, excluding participants based on data quality issues, transforming and constructing physiological variables, and modeling results. We encourage researchers to make these criteria and procedures explicit in preregistered analysis plans.
- Researchers should conduct sensitivity analyses to be sure that a particular stimulus or transformation strategy is not driving their results.
- Researchers should report all of the results from their data anlayses and make data available to other researchers.

Acknowledgments

We thank the National Science Foundation (SES-1423788) for generously supporting the workshops that led to the production of this manuscript.

References

Andreassi, J. (2007). *Psychophysiology*. New York: Psychology Press, https://doi.org/10.4324/9780203880340

Arceneaux, K., Dunaway, J., & Soroka,S. (2018). Elites are people, too: The effects of threat sensitivity on policymaker's spending priorities. *PLOS ONE*, 13(4), e0193781. https://doi.org/10.1371/journal.pone.0193781

Bakker, B. N., Schumacher, G., Gothreau, C., & Arceneaux, K. (in press). Conservatives and liberals have similar physiological responses to threats. *Nature Human Behaviour*.

Bakker, B. N., Schumacher, G., & Rooduijn, M. (2019, October 3). Hot politics: Affective responses to political rhetoric. OSF. https://osf.io/dqc74

Bernstein, A. S., Frith, C. D., Gruzeiler, J. H., Patterson, T., Straube, E., Venables, P. H., & Zahn, T. P. (1982). An analysis of the skin conductance orienting response in samples of American, British, and German schizophrenics. *Biological Psychology*, 14(3), 155–211. https://doi.org/10.1016/0301-0511(82)90001-1

Boucsein, W. (2012). *Electrodermal activity* (2nd ed.). Springer. https://doi.org/10.1007/978-1-4614-1126-0

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, 49(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Boudreau, C., McCubbins, M. D., & Coulson, S. (2009). Knowing when to trust others: An ERP study of decision making after receiving information from unknown people. *Social Cognitive and Affective Neuroscience*, 4(1), 23–34. https://doi.org/10.1093/scan/nsn034

Brader, T. (2006). *Campaigning for hearts and minds: How emotional appeals in political ads work*. University of Chicago Press.

Bradley, M. M., & Lang, P. J. (1999). *International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings* (Technical Report B-2). Center for Research in Psychophysiology, University of Florida.

Bradley, M. M., & Lang, P. J. (2007). *Affective norms for English text (ANET): Affective ratings of text and instruction manual* (Technical Report D-1). University of Florida.

Braithwaite, J. J., Watson, D. G., Jones, R. J., & Rowe, M. (2013). *A guide for analysing electrodermal activity (EDA) and skin conductance responses (SCRs) for psychological experiments* (Technical Report). Selective Attention and Awareness Laboratory, Behavioural Brain Sciences Centre, University of Birmingham, UK.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2007). *Handbook of psychophysiology* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511546396

Carlson, T. N., McClean, C. T., & Settle, J. E. (2020). Follow your heart: Could psychophysiology be associated with political discussion network homogeneity? *Political Psychology*, 41(1), 165–187. https://doi.org/10.1111/pops.12594

Clifford, S., & Jerit, J. (2014). Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Political Science*, 1(2), 120–131. https://doi.org/10.1017/xps.2014.5

Coppock, A., & Green, D. P. (2015). Assessing the correspondence between experimental results obtained in the lab and field: A review of recent social science research. *Political Science Research and Methods*, 3(1), 113–131. doi:10.1017/psrm.2014.10

Coppock, A., Leeper, T. J., & Mullinix, K. J. (2018). Generalizability of heterogeneous treatment effect estimates across samples. *Proceedings of the National Academy of Science of the United States of America*, 115(49), 12441–12446. https://doi.org/10.1073/pnas.1808083115

Cutmore, T. R. H., & James, D. A. (1999). Identifying and reducing noise in psychophysiological recordings. *International Journal of Psychophysiology*, 32(2), 129–150. https://doi.org/10.1016/S0167-8760(99)00014-8

Dawes, C. T., Loewen, P. J., Schreiber, D., Simmons, A. N., Flagan, T., McElreath, R., Bokemper, S. E., Fowler, J. H., & Paulus, M. P. (2012). Neural basis of egalitarian behavior.

Proceedings of the National Academy of Sciences, 109(17), 6479–6483. https://doi.org/10.1073/pnas.1118653109

Dawson, M. E., & Nuechterlein, K. H. (1984). Psychophysiological dysfunctions in the developmental course of schizophrenic disorders. *Schizophrenia Bulletin*, 10(2), 204–232. https://doi.org/10.1093/schbul/10.2.204

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. B. Berntson (Eds.), *Handbook of psychophysiology* (3nd ed., pp. 159–181). Cambridge University Press.

Druckman, J. N., & Kam, C. D. (2011). Students as experimental participants: A defense of the "narrow data base." In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), *Cambridge handbook of experimental political science* (pp. 41–57). Cambridge University Press. https://doi.org/10.1017/CBO9780511921452.004

Edelberg, R. (1967). Electrical properties of the skin. In C. C. Brown (Ed.), *Methods in psychophysiology* (pp. 1–53). Williams & Wilkins.

Fowles, D. C., Christie, M. J., Edelberg, R., Grings, W. W., Lykken, D. T., & Venables, P. H. (1981). Publication recommendations for electrodermal measurements. *Psychophysiology*, 18(3), 232–239. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), 1502–1505. https://doi.org/10.1126/science.1255484

Gerber, A., Arceneaux, K., Boudreau, C., Dowling, C., Hillygus, S., Palfrey, T., Biggers, D. R., & Hendry, D. J. (2014). Reporting guidelines for experimental research: A report from the Experimental Research Section Standards Committee. *Journal of Experimental Political Science*, 1(1), 81–98. https://doi.org/10.1017/xps.2014.11

Grossman, P., & Taylor, E. W. (2007). Toward understanding respiratory sinus arrhythmia: Relations to cardiac vagal tone, evolution and biobehavioral functions. *Biological Psychology*, 74 (2), 263–285. https://doi.org/10.1016/j.biopsycho.2005.11.014

Gruszczynski, M. W., Balzer, A., Jacobs, C. M., Smith, K. B., & Hibbing, J. R. (2013). The physiology of political participation. *Political Behavior*, 35(1), 135–152. https://doi.org/10.1007/s11109-012-9197-x

Hazlett, E. A., Dawson, M. E., Buchsbaum, M. S., & Nuechterlein, K. H. (1993). Reduced regional brain glucose metabolism assessed by positron emission tomography in electrodermal nonresponder schizophrenics: A pilot study. *Journal of Abnormal Psychology*, 102(1), 39–46. https://doi.org/10.1037/0021-843X.102.1.39

Hibbing, J. R., Smith, K. B., & Alford, J. R. (2014). Differences in negativity bias underlie variations in political ideology. *Behavioral and Brain Sciences*, 37(3), 297–350. https://doi.org/10.1017/S0140525X13001192

Journal Editors Transparency Statement (JETS). (2015). Data access and research transparency (DART). http://www.dartstatement.org/#!blank/c22sl

Klein, R. A., *et al.*, (2014). Data from investigating variation in replicability: A "many labs" replication project. *Journal of Open Psychology Data*, 2(1), e4. https://doi.org/10.5334/jopd.ad

Katsanis, J., & Iacono, W. G. (1994). Electrodermal activity and clinical status in chronic schizophrenia. *Journal of Abnormal Psychology*, 103(4), 777–783. https://doi.org/10.1037/0021-843X.103.4.777

Ksiazkiewicz, A., & Hedrick, J. (2013). An introduction to implicit attitudes in political science research. *PS: Political Science & Politics*, 46(3), 525–531. https://doi.org/10.1017/S1049096513000632

Lang, P. J. (1968). Fear reduction and fear behavior: Problems in treating a construct. In J. M. Shlien (Ed.), *Researchin psychotherapy* (p. 90–102). American Psychological Association. https://doi.org/10.1037/10546-004

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual* (Technical Report A-8). University of Florida.

LeDoux, J. (2015a). Afterword:Emotional construction in the brain. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 459–463). Guilford Press.

LeDoux, J. (2015b). Feelings: What are they & how does the brain make them? *Daedalus*, 144(1), 96–111. https://doi.org/10.1162/DAED_a_00319

Lodge, M., & Taber, C. S. (2013). *The rationalizing voter*. Cambridge University Press. https://doi.org/10.1017/CBO9781139032490

Lohani, M., Payne, B. R., & Strayer, D. L. (2019). A review of psychophysiological measures to assess cognitive states in real-world driving. *Frontiers in Human Neuroscience*, 13, 57. https://doi.org/10.3389/fnhum.2019.00057

MacDuffie, K., Knodt, A., Radtke, S., Strauman, T., & Hariri, A. (2019). Self-rated amygdala activity: An auto-biological index of affective distress. *Personality Neuroscience*, 2, E1. doi: 10.1017/pen.2019.1

Marcus, G. E., Neuman, W. R., & MacKuen, M. (2000). *Affective intelligence and political judgment*. University of Chicago Press.

Miguel, E., *et al.*, (2014). Promoting transparency in social science research. *Science*, 343(6166), 30–31. https://doi.org/.1126/science.1245317

Mutz, D. C., & Reeves, B. (2005). The new videomalaise: Effects of televised incivility on political trust. *American Political Science Review*, 99(1), 1–15. https://doi.org/10.1017/S0003055405051452

Osmundsen, M., Hendry, D., Laustsen, L., Smith, K., & Petersen, M. (2019, June 21). The Psychophysiology of Political Ideology: Replications, Reanalysis and Recommendations. https://doi.org/10.31234/osf.io/49hfg

Oxley, D. R., *et al.,* (2008). Political attitudes vary with physiological traits. *Science*, 321(5896), 1667–1670. https://doi.org/10.1126/science.1157627

Renshon, J., Lee, J., & Tingley, D. (2015). Physiological arousal and political beliefs. *Political Psychology*, 36(5), 569–585. https://doi.org/10.1111/pops.12173

Schaffer, F., Combatalade, D., Peper, E., & Meehan, Z. M. (2016). A guide to cleaner electrodermal activity measurements. *Biofeedback*, 44(2), 90–100. https://doi.org/10.5298/1081-5937-44.2.01

Smith, K. B., Oxley, D., Hibbing, M. V., Alford, J. R., & Hibbing, J. R. (2011). Disgust sensitivity and the neurophysiology of left-right political orientations. *PLOS ONE*, 6(10), e25552. https://doi.org/10.1371/journal.pone.0025552

Soroka, S. N. (2014). *Negativity in democratic politics: Causes and consequences*. Cambridge University Press. https://doi.org/10.1017/CBO9781107477971

Soroka, S. N. (2019). Skin conductance in the study of politics and communication. In G. Foster, (Ed.), *Biophysical measurement in experimental social science research: Theory and practice* (pp. 85–104). Academic Press.

Soroka, S., Fournier, P., & Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences*, 116 (38), 18888–18892. https://doi.org/10.1073/pnas.1908369116

Stern, R. M., Ray, W. J., & Quigley, K. S. (2001). *Psychophysiological recording*. Oxford University Press.

Tassinary, L. G., Cacioppo, J. T., & Vanman, E. J. (2007). The skeletomotor system: Surface electromyography. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (p. 267–299). Cambridge University Press. https://doi.org/10.1017/CBO9780511546396.012

van Dooren, M., & Janssen, J. H.2012. Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & Behavior*, 106(2), 298–304. https://doi.org/10.1016/j.physbeh.2012.01.020

Venables, P. H., & Mitchell, D. A. (1996). The effects of age, sex, and time of testing on skin conductance activity. *Biological Psychology*, 43(2), 87–101. https://doi.org/10.1016/0301-0511(96)05183-6

Vitriol, J. A., Larsen, E. G., & Ludeke, S. G. (2019). The generalizability of personality effects in politics. *European Journal of Personality*, 33(6), 631–641. https://doi.org/10.1002/per.2222

Wagner, M. W., Deppe, K., Jacobs, C. M., Friesen, A., Smith, K. B., & Hibbing, J. R. (2014). Beyond survey self-reports: Using physiology to tap political orientations. *International Journal of Public Opinion Research*, 27(3), 303–317. https://doi.org/10.1093/ijpor/edu036

Walker, F. R., Thomson, A., Pfingst, K., Vlemincx, E., Aidman, E., & Nalivaiko, E. (2019). Habituation of the electrodermal response—A biological correlate of resilience? *PLOS ONE*, 14(1), e0210078. https://doi.org/10.1371/journal.pone.0210078

Whitton, A. E., Henry, J. D., Rendell, P. G., & Grisham, J. R. (2014). Disgust, but not anger provocation, enhances levator labii superioris activity during exposure to moral transgressions. *Biological Psychology*, 96, 48–56. https://doi.org/10.1016/j.biopsycho.2013.11.012