P-curving A More Comprehensive Body of Research on Postural Feedback Reveals Clear Evidential Value For "Power Posing" Effects: Reply to Simmons and Simonsohn

Amy J. C. Cuddy
Harvard University, Department of Psychology

S. Jack Schultz
Harvard Business School

Nathan E. Fosse
Harvard University Institute for Quantitative Social Science (IQSS)

November 6, 2017

Forthcoming in *Psychological Science**

**Note: we will provide the final, copy-edited version of this article as soon as it is available.*

Several *p*-curving analyses based on a systematic review of the current scientific literature on the feedback effects of adopting expansive postures reveal strong evidential value for postural feedback effects (i.e., "power posing"), and particularly robust evidential value for effects on emotional and affective states (i.e., emotions, affect, mood, and evaluations, attitudes, and feelings about the self). These findings stand in contrast to those of Simmons and Simonsohn (2017), whose results from a *p*-curving analysis of an older and less comprehensive set of studies led them to conclude that the "existing evidence" does not possess evidential value.

Background

In a 2015 *Psychological Science* commentary, Carney, Cuddy, and Yap presented a narrative review of the psychological feedback effects of adopting expansive (vs. neutral or contractive) nonverbal postures (Here, we will refer to this as *postural feedback*.) While they had sought to include "all published tests (to [their] knowledge)," the aim of their commentary was to provide a theoretical and methodological summary of available experimental studies of the effects of postural feedback on various psychological outcomes, including cognitive, emotional, behavioral, and physiological measures; to compare and contrast similarities and differences across these studies and a conceptual replication attempt of the main study in Carney, Cuddy, & Yap (2010), which was reported in a 2015 *Psychological Science* commentary (Ranehill et al., 2015); to identify possible moderators of postural feedback effects; and to determine promising avenues for future research. It was neither intended to be nor was presented as a systematic review and meta-analysis of the literature.

In a 2017 *Psychological Science* commentary and in a 2015 Data Colada blog post, Simmons and Simonsohn (S&S) submitted the studies listed in the Carney et al. (2015) narrative review to a *p*-curving analysis, a meta-analytic technique described below, and concluded that their results do not support the existence of a real effect of "power posing" and that "the existing evidence is too weak to justify a search for moderators or to advocate for people to engage in power posing to better their lives."

Overview

We conducted a series of *p*-curve analyses, following Simonsohn et al.'s rules of *p*-curving (www.p-curve.com/guide.pdf) and using a systematically selected, comprehensive and updated set of published studies of "power posing," which yield starkly different results from those of S&S: evidential value for postural feedback across aggregated effects; evidential value for a clearly specified, theoretically important, single effect – feelings of power – which was omitted from the *p*-curve figures presented by S&S; and remarkably strong evidential value for a well-defined, theoretically important category of effects from the same set of studies identified in our systematic review -- all measures of "feelings," including emotions, affect, mood, and evaluations, attitudes, and feelings about the self.

We also discuss how it is that two groups of researchers, strictly applying the same analytic technique, can reach sharply conflicting conclusions about the extent to which an area of research does or does not contain evidential value.

P-curving

First, a brief primer on *p*-curving, a technique that was introduced by Simonsohn, Nelson, and Simmons in 2014: A "*p*-curve" is the distribution of statistically significant ($p < .05$) *p*-values selected for each study (one effect per study for the "main curve" and one effect per study for the "robustness curve") in a given set of studies that is defined by the "*p*-curver." Based on the distribution of *p*-values, the authors argue, one can "distinguish between sets of significant findings that are likely versus unlikely to be the result of selective reporting" (Simonsohn et al., 2014, p. 535), determining if the body of research constitutes "evidential value" (a right-skewed curve), "inadequate evidential value" (a flatter distribution than we would expect to find if the underlying studies had an average power of 33%, a threshold that S&S have described as arbitrary but justifiable), or "*p*-hacking" to achieve statistical significance (a left-skewed curve).
[1] According to Simonsohn et al. (2014), *p*-curving also produces an estimate of the average statistical power of the studies that corrects for selective reporting. Recently, Simonsohn et al. revised their methods, including not only tests of skew for all *p*s < .05, but also for *p*s < .025 (Simonsohn, Simmons, & Nelson, 2015).

All conclusions drawn from a *p*-curve analysis are necessarily constrained by the content of the input. In their 2017 comment and 2015 blog post, S&S restricted their analysis to the studies cited by Carney et al. (2015). Thus, their p-curve combined widely disparate dependent variables (e.g., pain tolerance, thought abstraction, self-reported vengeful intention, gambling, hormonal changes, eating behavior, to name a handful). Moreover, the S&S *p*-curve omitted many relevant studies. As already noted, they also excluded self-reported feelings of power as a DV. We report *p*-curve analyses that overcome those limitations, by systematically updating the set of studies and further examining clearly defined and theoretically meaningful subsets of effects. We believe that for *p*-curving to produce the most accurate and useful findings and conclusions, it must be applied using the best available evidence to test clearly specified a priori research hypotheses regarding well-defined effects, the aim of our analyses.

The Present Analyses

In the present analyses of the postural feedback literature, we aim to answer three meta-analytic questions that we defined a priori: First, does a systematic review of the literature pertaining to studies of the feedback effects of adopting expansive vs. contractive (or neutral) postural manipulations, consistent with standards established by *p*-curves, possess evidential value? Second, does the effect of postural feedback on a clearly specified, theoretically important single outcome, feelings of power, possess evidential value? Third, does the effect of postural feedback on a well-defined, theoretically meaningful and coherent category of findings from the

main curve, those measuring other emotional and affective states (e.g., emotions, affect, mood, and evaluations, attitudes, and feelings about the self) possess evidential value?

We began by conducting a systematic review of the literature with the aim of identifying the complete set of published empirical studies of "power posing" up to December 20, 2016. While narrative reviews provide a qualitative description of a body of literature (e.g., Carney et al., 2015), systematic reviews are based on a priori research questions regarding the evaluation of a body of theoretically relevant literature, which then guide careful and comprehensive study inclusion and exclusion (see, for example, Cooper, 2016; Uman, 2011). In our first *p*-curving analysis, our goal was to generate a *p*-curve based on our comprehensive search of the "power posing" literature and to compare it to the S&S curve. Any differences would provide information about whether and how the content of the included studies and effects can affect the results and conclusions. This analysis addresses our first methodological question: how do sample selection decisions influence the *p*-curving results and conclusions regarding this broad set of findings?

For our next analyses, acknowledging that limited conclusions can be drawn from these omnibus tests of aggregated effects, we refine the inputs to address our second methodological question: how does the undifferentiated aggregation of widely disparate effects into a single *p*-curve influence the conclusions that can be drawn about this broad set of findings?

Thus, in our second analysis, we *p*-curve one causal association between expansive posture and a clearly defined, theoretically meaningful single measure: the effect of postural expansiveness on *feelings of power*. As theorized by Carney et al. (2010; 2015) and by scores of social psychologists who study power (see Galinsky, Rucker, & Magee, 2015, for a review), feeling powerful is an intrinsically consequential, theoretically important, fundamental outcome. Feelings are core to the field's most popular self-definition: "Social psychology is the scientific study of how people's thoughts, feelings, and behaviors are influenced by the actual, imagined, or implied presence of others" (Allport, 1985, p. 3). As Wegner and Gilbert (2000, p. 1) explained, "...the center around which modern social psychology turns is the understanding of subjective experience… social psychology is intimately concerned with the scientific understanding of what it is like to be a person -- why our existence at this moment and in time and space *feels* the way it does" (emphasis not added). And it is a fait accompli that emotions and affect influence cognitive, behavioral, physiological, and other outcomes; this is, in fact, one of the key principles underlying much of social psychology in particular, and the social sciences in general (e.g., Bertrand & Mullainathan, 2001; Russell, 2003; Wegner & Gilbert, 2000). For example, subjective states and experiences, such as feelings of agency, happiness, and evaluations of the self, predict objective measures of behavior, health, and general wellbeing (e.g., Adler et al., 2015; Aneshensel, Phelan, & Bierman, 2013). And emotion theorists have long demonstrated the primacy of affect as preceding and motivating both cognition and behavior (see Zajonc, 1998, for a review).

Feeling powerful is an intrinsically consequential social psychological outcome that can influence impactful downstream outcomes. Specific to the psychology of power, hundreds of studies by researchers including Fiske, Galinsky, Guinote, Inesi, Keltner, Magee, Overbeck, P.

Smith, and others, have firmly established that subjective feelings of power influence both cognition, behavior, and physiological outcomes, including but not limited to stereotypes, resistance to influence, creativity and authenticity, physical and mental performance, self-regulation, goal pursuit, physiology, health, general wellbeing, and so on (e.g., Galinsky et al., 2015).

Moreover, theories of body-mind feedback investigating various effects of nonverbal behavior on people's emotional and affective states date back to William James's late eighteenth-century theories of emotion and ideomotor action (see Laird & Lacasse, 2014). Evidence that adopting postural expressions of emotions not only reflects, but also shapes, emotions, contributes to a foundational area of social psychological theory.

S&S excluded tests of effects of postural manipulations on self-reported feelings of power from their *p* curve analysis, on the ground that such measures are merely manipulation checks. We disagree. In the seminal Carney et al. (2010) article, self-reported power was repeatedly described as a DV of primary interest, from the abstract through to the discussion. The systematic review reported in the current article yielded 14 studies that measured feelings of power; 12 of those studies treated feelings of power as measures of theoretical interest. Only two studies, Cuddy, Wilmuth, Yap, and Carney (2015) and Ranehill et al. (2015), characterized self-reported feelings of power as a manipulation check, but we believe that was an error and certainly the vast majority of studies in this literature have not described feelings of power as a mere manipulation check. In fact, in some studies (e.g., Park et al., 2013, Studies 2a and 2b), feelings of power was the only DV, and was explicitly presented as the key outcome, not as a manipulation check (S&S excluded those studies from their main curve). So, it is not normative in this literature to treat feelings of power as a manipulation check.

It would simply not make sense to exclude feelings of power from a *p*-curve analysis of this literature. As theorized by Carney et al. (2010; 2015) and by scores of social psychologists who study power (e.g., Galinsky et al., 2015), feeling powerful is an intrinsically consequential, theoretically important outcome. Feelings are core to social psychology's most popular definition: "Social psychology is the scientific study of how people's thoughts, feelings, and behaviors are influenced by the actual, imagined, or implied presence of others" (Allport, 1985, p. 3). As Wegner and Gilbert (2000, p. 1) explained, "The center around which modern social psychology turns is the understanding of subjective experience… social psychology is intimately concerned with the scientific understanding of what it is like to be a person -- why our existence at this moment and in time and space feels the way it does" (emphasis in original).

Emotions and affect influence cognitive, behavioral, physiological, and other outcomes. This is one of the key principles underlying much of social psychology in particular, and the social sciences in general (e.g., Bertrand & Mullainathan, 2001; Russell, 2003; Wegner & Gilbert, 2000). It is, therefore, of central interest to determine if postural manipulations influence people's emotions and affect.

Our third and fourth *p*-curving analyses examine what happens when we reach beyond

feelings of power to look at the evidential value of postural feedback effects on other feelings – emotions, affect, mood, and evaluations, attitudes, and feelings about the self. Focusing on emotion and affect makes sense for several reasons. First, much of the research on postural feedback, which is theoretically grounded in the relationship between nonverbal expressions and emotion, has naturally focused on the effects of expansive postures on emotional and affective states (as opposed to cognition and behavior); it is of primary theoretical interest. Second, it allows us to address concerns about undifferentiated aggregation without limiting the analysis to a single emotion, feelings of power. Third, by including the entire set of emotion- and affect-related outcomes while excluding feelings of power, we can confront questions about whether postural feedback effects are merely demand effects. The remaining set of emotion- and affect-related outcomes includes findings obtained in procedures that seem unlikely to be susceptible to demand characteristics. In some such studies there were no obvious cues as to what sort of response was "demanded" (e.g., mood recovery, changes in various discrete emotion states, changes in negative affect, open-ended thoughts-listing task followed by assignment of valence to each), and/or the outcome variables seem difficult to control or to "fake" (e.g., speed of retrieval of positive and negative personal memories, mood recovery, ability to recall positive vs. negative words from a list presented earlier in the study, changes to discrete emotion states embedded in a long list of emotions).

Applying a systematic coding procedure, we limit inclusion in the third curve to all emotion- and affect-related outcomes, which we refer to as EASE (Emotions, Affect, and Self-Evaluation) variables, while excluding feelings of power. Excluding feelings of power from our assessment of evidential valued for postural feedback on emotion-related effects makes that analysis more conservative.

While EASE variables represent a theoretically meaningful subset of the effects included in our first analysis of aggregated outcomes, non-EASE measures do not; they are the conceptually heterogeneous effects that remain after extracting the theoretically coherent set of EASE variables. Additionally, conclusions based on *p*-curve analyses of non-EASE variables require the same caution required for the theoretically heterogeneous set of measures involved in the omnibus *p*-curve analyses by S&S (2017). We describe our categorization methods for EASE and non-EASE variables in further detail below.

Analytic Approach

Taking a conservative analytic approach, we *p*-curved the postural feedback literature applying the selection criteria and statistical methods prescribed by S&S, applying only two distinct differences in analytic approach: First, we defined our questions a priori and systematically gathered all available data relevant to the questions at hand. Second, as described above, we conducted several *p*-curve analyses – one for the aggregated outcomes, as S&S did, one for feelings of power, and a pair for EASE and non-EASE variables, respectively. We present all *p*-curve analysis results below.

In addition to the original 20 papers (34 studies) assessed by Carney et al., 2015, we performed a systematic literature search for additional studies. We searched for peer-reviewed studies using the Harvard Hollis+ platform. From the known literature, we identified the following keywords (e.g. “power,” “dominance,” “pride,” “proud,” “guilt,” “shame,” “mood,” “emotion,” “expansive,” “open,” “upright,” “contractive,” “slouched,” “slumped,” “hunched,” “closed,” “pose,” “posing,” “posture,” and "incidental posture") resulting in numerous queries.

All studies had to feature a postural manipulation that (directly or indirectly) induced expansive or contractive nonverbal postures, consistent with prior definitions (Carney et al., 2010, 2015). We included only studies with postural manipulations that involved a modification of the orientation/openness of the chest (or torso), and/or shoulders; studies that manipulated only head orientation (i.e., chin and head down vs. chin and head upright), for example, were not included. Postural manipulations could also include changes to the orientation of the arms, legs, head and neck, though each of these elements was not considered sufficient on its own. Because postural expansiveness is a continuous spectrum, one of the posture conditions had to be more expansive relative to the other(s). Studies that featured sitting or standing postural effects were included while any studies testing the effects of supine postures or movement (i.e. walking or dancing) were excluded. (For additional information about our systematic literature review methods, please see the supplementary materials on OSF.)

The literature search produced an additional 21 studies that met all criteria for inclusion. Those 21 studies, added to the original 34 studies, resulted in a sample of 55 studies. All 55 studies meet our inclusion criteria and all 55 studies are accounted for in our systematic literature search results. From each study, we selected the appropriate statistic(s) based on the rules provided by the *p*-curve guide and the selection criteria used by S&S. We selected the first reported hypothesis pertaining to postural feedback effects when the hypothesis was clearly stated. For our robustness curve, we followed the practices used by S&S in (1) carrying over main results where the *p*-curve guide did not require a specific alternative statistic, (2) selecting the specified alternative statistic in cases where the *p*-curve guide required it, and (3) including the appropriate statistic for a second hypothesized effect in cases where there were multiple hypothesized effects. When the article presented multiple hypothesized effects pertaining to postural feedback, we included the second statistic that was explicitly hypothesized (e.g. “we hypothesized that expansive posture would increase testosterone and decrease cortisol”) or the second reported statistic pertaining to a general hypothesis (e.g. “we hypothesized that posture would impact hormones”). (See the disclosure table in our supplementary materials for detailed information regarding all statistics that were selected and included in each *p*-curve.)

The EASE *p*-curve required reliable categorization of variables as EASE or non-EASE. Five experts coded the variables: the first two authors of this paper and, to insure objectivity, three additional expert coders, all of whom are social psychologists and tenured professors at research universities but none of whom does research on postural feedback. Coders were contacted by email that included a link to an online survey. They were asked to categorize all measures that

were drawn from our systematically selected set of peer-reviewed experiments testing postural feedback effects and that were included in the omnibus curve. Coders were provided with the list of dependent measures, named as they were by the original researchers, along with excerpts from the original articles that described exactly how the variables were operationalized. Including both the names and operationalizations of the variables insured that the coding was indeed based on what was actually measured, given that there are sometimes discrepancies between the conceptual variable and the operationalized variable. Coders were asked to identify, "measures of emotions, affect, mood, and evaluations, attitudes, and feelings about the self (i.e., self-evaluations)," adding that "measures can be explicit or implicit, direct or indirect, but they should be primary measures, as opposed to correlates, of an EASE construct." The intercoder correlation coefficient (ICC), which we calculated to assess inter-rater agreement, indicated excellent reliability[2], *ICC average* = .92 (*CI*: 88-.95). The final categorizations of variables as EASE or non-EASE were determined by majority rule. The final EASE set included such variables as retrieval of positive and negative memories, mood recovery, changes in specific emotion states, recall of positive and negative words from lists presented earlier in the study, and self-evaluations, demonstrating that the effects of postural feedback on affective variables clearly extend beyond feeling powerful. (For more details about the coding of EASE and non-EASE variables, refer to the disclosure table our supplementary materials on OSF.)

Results

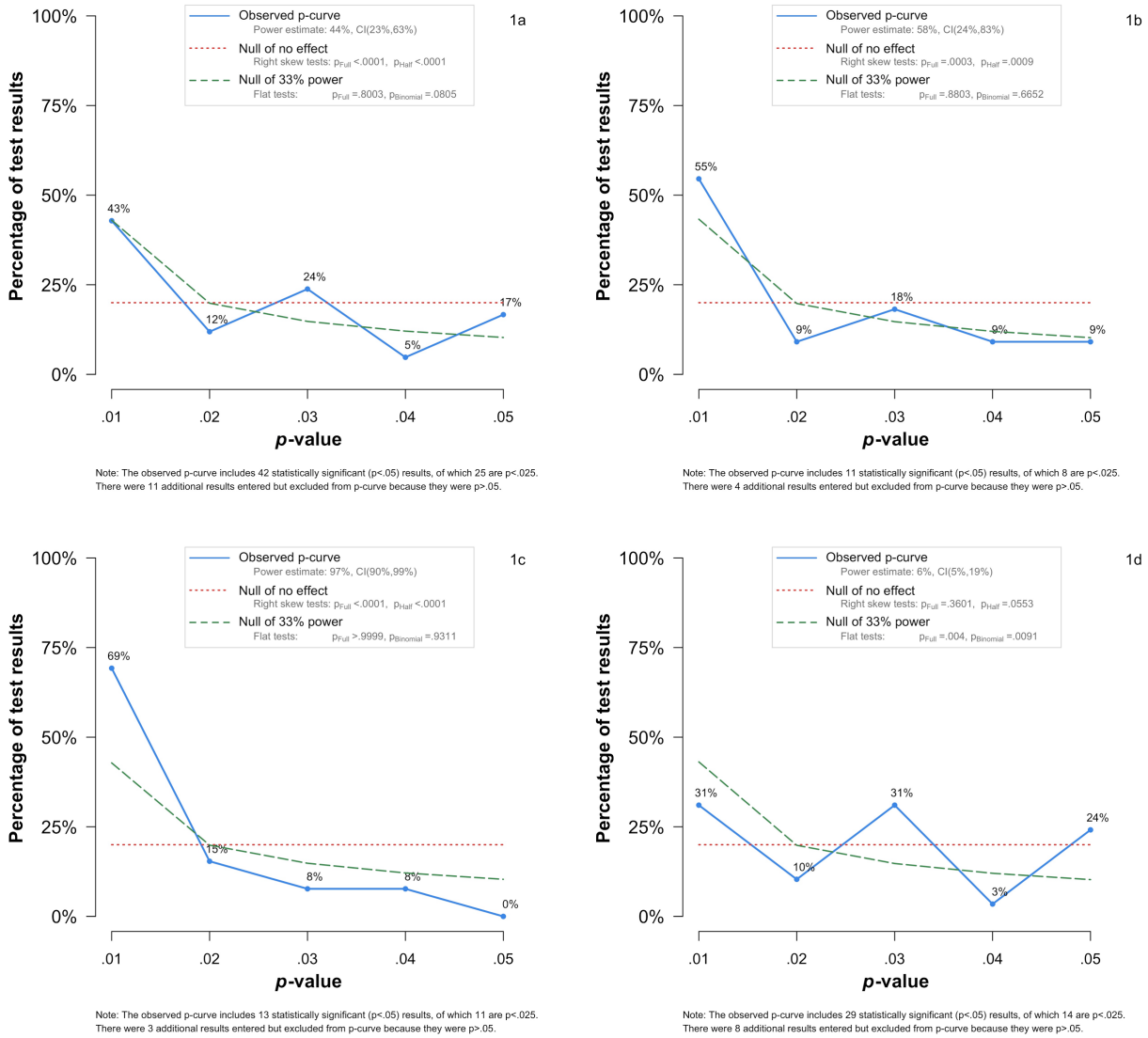
Omnibus Test: Evidential Value of Postural Feedback Effects on Aggregated Variables

Our first *p*-curve analysis, based on a systematic literature review that aimed to include all published empirical tests of "power posing" manipulations as of December 20, 2016, comprises 53 statistical results and clearly demonstrates that the postural feedback literature contains strong evidential value (Figure 1a). This *p*-curve serves as the comparator to the main curve presented by S&S. In fact, we found evidential value both in our "main" and "robustness" *p*-curves, as well as with the half *p*-curve featured in the latest version of the *p*-curve (version 4.05), which assesses evidential value among studies with *p*'s below the median.[3]

A literature is determined to "[contain] evidential value if either the half *p*-curve … is significantly right-skewed at the 5% level or if *both* the half *p*-curve and the full *p*-curve are significantly right-skewed at the 10% level" (Simonsohn et al., 2015). In this case, all conditions for evidential value are met, showing clear right-skew for both p_{half} (< .0001) and p_{full} (< .0001) in our main *p*-curve, *and* for both p_{half} (< .0001) and p_{full} (< .0001) in our robustness *p*-curve. Second, the observed *p*-curve is compared to "what is expected when studies have an average power of only 33%" (Simonsohn et al., 2015); a $p < .05$ for the full *p*-curve (or $p < .10$ for the full *and* binomial *p*-curve) would indicate a flatter curve than we would expect when the included studies have an average power of 33% and an absence of evidential value. The results of this analysis do *not* meet any of the criteria for an *absence* of evidential value, $p_{full} = .8003$, $p_{binomial} = .0805$ for our main *p*-curve; and $p_{full} = .9036$, $p_{binomial} = .1184$ for our robustness curve. Third, we find that the estimate of

average power for the set of studies is 44% in our main curve and 49% in our robustness curve (compared to the S&S estimates of 5% in both curves). When submitting this systematically identified current set of studies to *p*-curving analysis, the results demonstrate that the literature on postural feedback possesses evidential value.

Figure 1.
P-Curves of Postural Feedback Literature: Aggregated Effects, Feelings of Power, EASE Variables, non-EASE Variables



Note: 1a presents the main p-curve for aggregated postural feedback effects; 1b presents the p-curve for feelings of power; 1c presents the p-curve for EASE (Emotions, Affective attitudes, Self-evaluations) variables; 1d presents the p-curve for non-EASE variables.

In sum, the results of our omnibus *p*-curve, which included a systematically identified and comprehensive set of studies, demonstrate that this literature possesses evidential value. This finding sharply diverges from the results and conclusions of S&S's *p*-curve analysis, which failed to show evidential value. (Although note that no *p*-curving analysis by either set of authors yielded results that were left-skewed or that suggested the existing evidence was "*p*-hacked.")

Evidential Value for Postural Feedback Effects on Feelings of Power

In Figure 1b, we present the *p*-curving analysis for postural feedback effects on feelings of power, which clearly demonstrates that the effect of expansive posture on feelings of power possesses evidential value. First, the analysis yields strong evidence of right-skew for both p_{half} (.0009), and p_{full} (< .0003)[4]. Second, the results of the tests for flatness do not meet any of the criteria for an absence of evidential value, $p_{full} = .8803$, $p_{binomial} = .6652$. Third, the estimated average power for the specific feelings of power outcome is 58%, higher than the estimated average power for our omnibus curve.

This *p*-curving analysis shows strong evidential value for postural feedback effects on feelings of power -- a clearly specified and theoretically important single outcome.

Evidential Value for Postural Feedback Effects on EASE (Emotion, Affect, & Self-Evaluation) and Non-EASE Variables

EASE Variables. In Figure 1c, we present the *p*-curving analysis for postural feedback effects on EASE variables, which clearly reveals robust evidential value for postural feedback effects on emotion, affect, and self-evaluation (EASE) outcomes. First, the analysis yields very strong evidence of right-skew, $p_{half} < .0001$, $p_{full} < .0001$. Second, the results of the tests for flatness do not meet any of the criteria for an absence of evidential value, $p_{full} > .9999$, $p_{binomial} = .9311$. Third, the estimated average power for the EASE variables is extremely high, 97%, well exceeding the estimated average power of both the omnibus and feelings of power curves.

This *p*-curving analysis of a well-defined, theoretically important category of postural feedback effects - measures of emotions, affect, and self-evaluations -- demonstrates very strong evidential value. Expansive vs. contractive posture affects not only how powerful people feel, but how people feel on a wide variety of other emotion- and affect-related outcomes.

Non-EASE Variables. In Figure 1d, we present the *p*-curving analysis for postural feedback effects on non-EASE variables, a miscellaneous subset of the statistics featured in the main curve (Figure 1a) of our omnibus test. These are the theoretically heterogeneous "leftovers" after extracting the EASE effects. The test for right-skew is marginally significant for the half curve ($p_{half} = .0553$) and nonsignificant for the full curve ($p_{full} = .3601$). Additionally, tests for a null of 33% power indicate an *absence* of evidential value, $p_{full} = .0040$, $p_{binomial} = .0091$).

Although a significant half *p*-curve would be adequate to determine that a set of studies possesses evidential value (Simonsohn et al., 2015), the flatness tests clearly fail to reject the null

of 33% power. Thus, for this nebulous set of non-EASE variables, the *p*-curving analysis yields very weak support for the existence of evidential value.

Discussion

These analyses bring to light several critical discoveries about the existing postural feedback literature. When including a more comprehensive and current set of evidence, comprising 55 studies identified through a systematic review, *p*-curve analyses reveal (1a) clear evidential value for postural feedback on an aggregated set of effects, (1b) strong evidential value for a clearly specified and theoretically important single outcome, feelings of power, (1c) very strong evidential value for a well-defined and theoretically important category of other "feelings" effects -- emotions, affect and self-evaluations (i.e., EASE variables, which did not include feelings of power), and (1d) an absence of evidential value for the conceptually heterogeneous non-EASE effects that remained after separating out the EASE variables. Our findings also suggest that *p*-curving is likely to yield more accurate and informative results when researchers address the following practices: (2a) faulty sample selection decisions, and (2b) undifferentiated aggregation of disparate effects. When these practices are not adequately addressed, *p*-curve conclusions can lead to misguided dismissals of broad areas of research.

Strong Evidential Value for Postural Feedback Effects, Particularly for Emotions

Our *p*-curve analyses of emotion- and affect-related outcomes yielded robust evidence that postural feedback influences self-reported affective states. First, we found strong evidential value for a precisely specified outcome, feelings of power. That finding converges with a recent Bayesian meta-analysis of a new set of studies that, as described by Cesario, Jonas, & Carney (2017), "showed a reliable non-zero effect on felt power."[5] Presenting the results, Gronau et al. (2017) write, "Our meta-analysis yields very strong evidence for an effect of power posing on felt power." In the set of studies presented in our analyses, 11 studies demonstrated a significant effect of power posing on feelings of power; that does not include studies from 2017, which would increase the total number of replications. Together, the collective evidence provides strong support for the effect of postural feedback on feelings of power. From our theoretical perspective, an expansive posture is a universal expression of power and adopting such a pose leads people to feel more powerful. The finding of evidential value for self-reported feelings of power directly supports that claim. Moreover, we believe that even transient feelings of power can have long-lasting consequences for people's lives (e.g., Galinsky et al., 2015).

The robust evidential value for postural feedback effects on EASE variables – emotions, affect, and self-evaluations -- is particularly illuminating. These findings from the present set of studies provide convincing evidence that postural manipulations affected subjects' specific emotions, affect, mood recovery, retrieval and recall of positive vs. negative memories, and self-evaluations, demonstrating that the effects of postural feedback on affective variables clearly extend beyond causing people to feel more powerful. It is worth noting that the direction of most

of the EASE effects are consistent with Keltner, Gruenfeld, and Anderson's (2003) approach/inhibition theory of power: power activates the behavioral approach system (BAS; e.g., recall of more positive than negative words from memory, improved general mood and mood recovery, increased feelings of strength, decreased feelings of fear, for example).

Many studies that are featured in our EASE curve were likely robust to potential demand characteristics, by using a single or double-blind study design, deception, "non-deceptive obfuscation" (Zizzo, 2010), testing hypotheses that were simply not intuitive to participants (e.g., mood recovery, changes in various discrete emotion states, changes in negative affect, open-ended thoughts-listing task followed by assignment of valence to each), or by directly tracking the extent to which participants guessed the hypothesis in exit interviews (showing that they did not). Some studies were more resilient to demand effects because responses were implicit or otherwise difficult for participants to control (e.g., speed of retrieval of positive and negative personal memories, mood recovery, ability to recall positive and negative words from a list presented earlier in the study, changes to discrete emotion states embedded in a long list of emotions), responses were embedded in a broader survey instrument (e.g., changes in discrete self-reported emotions embedded in a long list of emotions), or, as demonstrated in recent research on demand effects in survey research, participants likely varied in their orientation such that some would have wanted to confirm the hypothesis, some to disconfirm it, and others would have been indifferent (Mummolo & Peterson, 2017). Citations for each of these examples are listed in our supplemental materials. Our assessment of the input for our EASE p-curve analysis, the strongest *p*-curve presented, is that it is unlikely that these postural feedback effects are demand effects, given the study designs and the latest research on demand characteristics.

In contrast to the EASE *p*-curve, the non-EASE *p*-curve comprises a theoretically heterogeneous, non-cohesive collection of effects (e.g., number of calories consumed at a meal, pain threshold, vengeful intentions, performance on creativity tasks, hormones, beliefs about religion, performance in a job interview, gambling, etc.), making any results, whether they indicate a presence or absence of evidential value, difficult to interpret. Removing the EASE variables "flattens" the curve for the remaining effects, which could indicate that evidential value for behaviors and hormones is weak. This interpretation is consistent with the mostly null results of the set of studies in the recent special issue of *CRSP* (Jonas et al., 2017) that measured effects of power posing on various behavioral outcomes. However, many of the non-EASE effects include non-behavioral or hormonal effects like cognitive abilities, creativity, and attitudes; the evidence for these effects seems to be stronger. It's also worth noting that the set of non-EASE effects include measures that are susceptible to demand characteristics, such as gambling, pain tolerance, and action tendencies in hypothetical scenarios. There is also a need for experimental tests of incremental or longitudinal effects of adopting expansive postures over time on various outcomes. Right now, we are not aware of any such research. As more studies are conducted and published, it will become easier for researchers to analyze other theoretically meaningful subsets of effects, such as hormonal effects, performance under stress, risk preferences, and cognitive abilities. Such

analyses of these subsets will continue to enhance the definition of this picture.

What do these analyses tell us about the evidence for postural feedback, or power posing, effects? Based on the present *p*-curving analyses, as strictly interpreted in accordance with the rules of *p*-curving, one must first conclude that the current literature on postural feedback does possess evidential value. By systematically identifying and analyzing meaningful subsets of effects, *p*-curving begins to give more definition to our findings and to the overall picture: the existing evidence of effects of postural feedback on feelings possess extremely strong evidential value. As the overall body of studies grows, it will become easier to analyze other meaningful subsets, like cognitive measures, performance behaviors, and psychophysiological outcomes. Combining these more focused meta-analyses of meaningful categories of effects with new, theory-driven studies that employ improved methods (e.g., pre-registration of a priori hypotheses, larger samples, more accurate hormone measurement tools) and that come from various disciplines will advance and refine our theoretical understanding of postural feedback – and the same will be true other areas of research -- leading to the identification of contextual variables that moderate effects and helping us to resolve conflicting evidence from studies of some of the specific effects, such as hormones and risk-taking, which have produced both significant and null effects. The analyses do not tell us, however, about the extent to which there is evidential value for other meaningful categories of effects, which individual postural feedback effects are most robust, which of them might be false positives, and how these complex relationships among posture and these many different variables may be affected by various moderators. It should go without saying that these curves are not the 'final' curves. No meta-analysis can be the final meta-analysis, because results hinge entirely on the content of what is included, and that content will continue to grow and change. Science is cumulative by nature.

How did two groups of researchers reach such discrepant findings and conclusions about the same area of research? Our analyses reveal two of the practices that contributed. First, we addressed the issue of *sample selection decisions* that may lead to an incomplete or non-representative set of studies and/or effects for inclusion in the analysis. Differences between S&S's selections and our selections gravely influenced the results of the S&S analysis and the conclusions they drew from those results, which dramatically differed from the results and conclusions from our analyses that were guided by our a priori systematic review of the entire body of literature. As Simonsohn et al. (2014) wrote in their seminal paper, "For inferences from *p*-curve to be valid, studies and *p*-values must be appropriately selected." Please note that sample selection is not limited to the selection of studies; it can also extend to the selection of effects from each study, particularly when a study includes multiple DVs that are equally weighted by the primary researchers. Second, our *p*-curve analyses of feelings of power and EASE variables underscore our concerns that *undifferentiated aggregation* can muddy the waters, making it difficult to draw accurate conclusions from *p*-curve analyses of widely disparate effects. In the present case, the results from S&S (2017) mask markedly strong effects of postural expansiveness on feelings of power and on other emotional and affective states.

We are not arguing that the statistical results of S&S's *p*-curving analysis are incorrect; we are arguing that their results and conclusions, as a result of the practices described above, are misleading with regard to assessments of the evidential value of this area of research. The present *p*-curving results annul S&S's conclusion that "the existing evidence is too weak to justify a search for moderators or to advocate for people to engage in power posing to better their lives." Our findings, including modest support for the general literature on postural feedback and markedly strong support for effects on emotional and affective states, should encourage researchers who are conducting research in this area, to continue doing so.

Acknowledgements

The authors thank the following people for providing constructive, expert input at various times and in various ways to the development of this paper: Xiang Ao, Don Carlston, Paul Coster, Naomi Ellemers, Eli Finkel, Susan Fiske, Dan Gilbert, Gary King, Andrew Marder, Allen McConnell, and Terri Vescio.

References

Allport, A. (1985). The historical background of social psychology. In G. Lindzey & E. Aronson (Eds.). Handbook of social psychology (Vol. 1, 3rd ed., pp. 1-46). New York: Random House.

Aneshensel, C. S., Phelan, J. C., & Bierman, A. (2013). The sociology of mental health: Surveying the field. In *Handbook of the sociology of mental health* (pp. 1-19). Springer Netherlands.

Bertrand, M., & Mullainathan, S. (2001). Do people mean what they say? Implications for subjective survey data. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=260131

Carney, D. R., Cuddy, A. J. C., & Yap, A. J. (2010). Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance. *Psychological Science*, *21*(10), 1363–1368. http://doi.org/10.1177/0956797610383437

Carney, Cuddy, A. J. C., & Yap, A. J. (2015). Review and Summary of Research on the Embodied Effects of Expansive (vs. Contractive) Nonverbal Displays. *Psychological Science*, *26*, 657–663. http://doi.org/10.1177/0956797614566855

Cesario, J., Jonas, K. J., & Carney, D. R. (2017). CRSP special issue on power poses: what was the point and what did we learn? *Comprehensive Results in Social Psychology, 2*, 1-5.

Cooper, H. (2016). *Research Synthesis and Meta-Analysis: A Step-by-Step Approach*. SAGE Publications.

Cuddy, A. J. C., Wilmuth, C. A., Yap, A. J., & Carney, D. R. (2015). Preparatory power posing affects nonverbal presence and job interview performance. *Journal of Applied Psychology*, *100*, 1286–1295. http://doi.org/10.1037/a0038543

Galinsky, A. D., Rucker, D. D., & Magee, J. C. (2015). Power: Past Findings, Present Considerations, and Future Directions. In M. Mikulincer, & P. R. Shaver (Eds.), *APA Handbook of Personality and Social Psychology* (Vol. 3: Interpersonal Relationships, pp. 421-460). Washington, DC: American Psychological Association.

Gronau, Q.F., Van Erp, S., Heck, D.W., Cesario, J., Jonas, K.J., & Wagenmakers, E-J. (2017). A Bayesian model-averaged meta-analysis of the power pose effect with informed and default priors: the case of felt power. *Comprehensive Results in Social Psychology*.

James, W. (1994). The physical basis of emotion. *Psychological Review, 101*, 205-210. (Original work published 1896).

Jonas, K., Cesario, J., Alger, M., Bailey, A.H., Bombari, D., Carney, D., Dovidio, J.F., Duffy, S., Harder, J., van Huistee, D., Jackson, B., Johnson, D.J., Keller, V.N., Klaschinski, L., LaBelle, O., LaFrance, M., Latu, I.M., Morssinkhoff, M., Nault, K., Pardal, V., Pulfrey, C., Rohleder, N., Ronay, R., Smart Richman, L., Schmid Mast, M., Schnabel, K., Schröder-Abé, M., & Tybur, J.M. (2017). Power poses – where do we stand? *Comprehensive Results in Social Psychology, 2,* 139-141.

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological review*, *110*, 265.

Laird, J. D., & Lacasse, K. (2014). Bodily influences on emotional feelings: Accumulating evidence and extensions of William James's theory of emotion. *Emotion Review, 6*, 27–34.

Mummolo, J., & Peterson, E. (2017). Demand Effects in Survey Experiments: An Empirical Assessment. Retrieved from https://papers.ssrn.com/abstract_id=2956147

Niedenthal, P. M. (2007). Embodying emotion. *Science, 316*, 1002–1005.

Ranehill, E., Dreber, A., Johannesson, M., Leiberg, S., Sul, S., & Weber, R. A. (2015). Assessing the Robustness of Power Posing No Effect on Hormones and Risk Tolerance in a Large Sample of Men and Women. *Psychological Science, 26,* 653-656.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145-172.

Simmons, J. P., & Simonsohn, U. (2016). *Power Posing: p-Curving the Evidence* (SSRN Scholarly Paper No. ID 2791272). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2791272

Simmons, J. P., & Simohnson, U. (2015, May 8). [37] Power Posing: Reassessing The Evidence Behind The Most Popular TED Talk - Data Colada. Retrieved from http://web.archive.org/web/20160613100317/http://datacolada.org/37

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). *P*-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*, 534.

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better *p*-curves: Making *p*-curve analysis more robust to errors, fraud, and ambitious *p*-hacking, a Reply to Ulrich and Miller (2015). *Journal of Experimental Psychology: General*, *144*, 1146–1152. http://doi.org/10.1037/xge0000104

Uman, L. S. (2011). Systematic reviews and meta-analyses. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *20*, 57.

Wegner D.M., Gilbert, D.T. (2000). Social psychology–the science of human experience. In H. Bless & J. Forgas (Eds). Subjective experience in social cognition and behavior (pp. 1-9). Philadelphia: Psychology Press.

Zajonc, R. B. (1998). Emotions. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), Handbook of Social Psychology (4th ed., Vol. 1, pp. 591–634). Boston: McGraw-Hill.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics, 13*, 75-98.

---

[1] We also applied an earlier version of the *p*-curve app that included output on the likelihood of "*p*-hacking" as indicated by left-skew. There was no evidence of *p*-hacking in any analysis of any of our or S&S's *p*-curves. The R script we used (written by Uri Simonsohn) is available online at www.p-curve.com. Earlier versions of the *p*-curve app R scripts are presently unavailable though *p*-curve app updates are listed at http://www.p-curve.com/app4/versions.php. We include a copy of the *p*-curve app version 4.05 along with our disclosure table that are available here: https://osf.io/pfh6r/

[2] For guidelines, see https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4913118/

[3]The lowest *p*-value featured in our main curve corresponds to a measure of self-reported strength, drawn from the

study by Peper et al. (2016) in which participants held an erect or collapsed posture and resisted downward pressure applied to their arms and reported how strong they felt. The study protocol was designed to minimize demand characteristics, and the results show a clear effect on how posture affects felt strength.

[4] Following the *p*-curving practice of carrying over the main results where the *p*-curve guide does not require a specific alternative statistic, the feelings of power robustness curve does not have any alternative statistics to draw from, so it is not possible to produce a robustness curve that is not entirely redundant with the main curve.

[5] Our systematic review end date was December 20, 2016, and we judged it inappropriate to add studies that we learned of incidentally thereafter because doing so would have undermined the objectivity and integrity of the systematic review. Consequently, these and other 2017 studies were not included in our set of studies (although we provide references to them in our supplemental materials on the OSF).