COGNITION AND EMOTION
0000, 00 (00), 1–28

Psychology Press
Taylor & Francis Group

The evaluative space grid: A single-item measure of positivity and negativity

Jeff T. Larsen
Texas Tech University, Lubbock, TX, USA

Catherine J. Norris
Dartmouth College, Hanover, NH, USA

A. Peter McGraw
University of Colorado, Boulder, CO, USA

Louise C. Hawkley and John T. Cacioppo
University of Chicago, Chicago, IL, USA

The authors introduce the *evaluative space grid* (ESG), a two-dimensional grid that provides a single-item measure of positivity and negativity. In Study 1, ESG ratings of gamble outcomes were highly correlated with those obtained from conventional, less-efficient, unipolar measures, thus providing evidence for the grid's convergent validity. In Study 2, participants rated their moment-by-moment evaluative reactions to gamble outcomes with the grid every 100 ms; results replicated earlier findings that some outcomes elicit only positivity or negativity whereas others simultaneously elicit positivity and negativity. In Studies 3 and 4, the difference between the grid's positive and negative ratings of several types of stimuli and bipolar valence ratings were highly correlated, thus demonstrating the grid's generalisability and predictive validity. Study 4 also showed that ESG ratings predicted facial electromyographic activity, particularly in tasks involving strongly affective stimuli. Taken together, results indicate that the grid provides efficient, valid indices of positivity and negativity.

Would an employee who anticipated a $1200 raise but received only a $500 raise feel pleased about the raise or displeased because it fell well short of expectations (Kahneman, 1992)? The structure of the question presupposes

---

Correspondence should be addressed to: Jeff T. Larsen, Department of Psychology, Texas Tech University, Lubbock, TX 79409–2051, USA. E-mail: jeff.larsen@ttu.edu

This research was supported by NIMH Grant P50 MH52384–01A1.

The authors thank Barbara Mellers for making available the data from Study 2 and Stephen Crites, Jonathan Levav, and Darcy Reich for helpful comments on previous versions.

www.psypress.com/cogemotion DOI: 10.1080/02699930801994054

that evaluative reactions to such *disappointing wins* (Larsen, McGraw, Mellers, & Cacioppo, 2004) fall along a bipolar valence dimension ranging from good to bad and, as a result, precludes consideration of whether the employee would feel both pleased *and* displeased. For decades, however, attitudes theorists have used unipolar measures of positivity and negativity to demonstrate that people can feel both good and bad (i.e., ambivalent) about such attitude objects as capital punishment and abortion (e.g., Kaplan, 1972; Scott, 1968). Similarly, based on evidence that some positive and negative moods and emotions are uncorrelated in experience (e.g., Bradburn, 1969; Diener & Emmons, 1984; Watson & Tellegen, 1985) emotion researchers have used unipolar scales to investigate the structure of emotion (e.g., Watson, Clark, & Tellegen, 1988). Nevertheless, contemporary unipolar measures of positivity and negativity can require participants to make numerous ratings. The purpose of this article is to introduce and examine the validity of a single-item measure of positivity and negativity.

The evaluative space model

Building on the pioneering work of Scott (1968) and Bradburn (1969), Cacioppo and Berntson's (1994) evaluative space model (ESM) contended that evaluative processes represent the integration of two separable and partially distinct components of the evaluative system. Whereas positivity is attuned to appetition and fosters approach, negativity is attuned to threat and fosters avoidance. One implication of the ESM is that the same stimulus can evoke ambivalent positive and negative reactions. Bipolar scales, however, can allow ambivalence to masquerade as neutrality by preventing respondents from reporting that they feel both good and bad. Indeed, Mellers, Schwartz, Ho, and Ritov (1997) found that disappointing wins are assigned relatively middling ratings on bipolar scales, but this does not imply that disappointing wins are evaluated as neutral. In fact, Mellers et al.'s findings raise the possibility that disappointing wins elicit both positivity *and* negativity. Consistent with this hypothesis, Larsen et al. (2004) found that participants felt more ambivalent about disappointing wins (e.g., winning $5 instead of $12) than outright wins (e.g., winning $5 instead of $3). Similarly, participants felt more ambivalent about *relieving losses* (e.g., losing $5 instead of $12) than outright losses (e.g., losing $5 instead of $3).

Measures of positivity and negativity

Attitudes researchers have used simple unipolar scales comprising one or more items intended to assess positivity (e.g., "How good do you feel about capital punishment?") and an additional one or more items intended to assess negativity (e.g., "How bad do you feel about capital punishment?").

With the bivariate evaluations and ambivalence measures (BEAMS; Cacioppo, Gardner, & Berntson, 1997), for instance, participants rate attitude objects in terms of positive adjectives (e.g., good, supporting, satisfying) and negative adjectives (e.g., unfavourable, unpleasant, disapproving). The BEAMS has been useful in studies of impression formation (Cacioppo et al., 1997) and evaluative reactions to pictures (Ito, Cacioppo, & Lang, 1998). By virtue of being multi-item scales, however, the BEAMS and all other existing unipolar measures of evaluative reactions can be time consuming when participants are asked to judge large numbers of stimuli.

Emotion researchers have also used simple unipolar scales to measure happiness, sadness, and other discrete emotions (e.g., Watson et al., 1988). Based on evidence that people often mistake unipolar measures of happiness for bipolar measures of happiness and sadness, however, Russell and Carroll (1999) advocated the use of more complex, dichotomous-then-unipolar scales first used by Ekman et al. (1987) and Reisenzein (1995). To report happiness, participants are first asked, "Do you feel happy?" and only those who check "yes" are asked to indicate how happy they felt on a 6-point scale. Similar dichotomous-then-unipolar items are used to measure sadness and other emotions.

Individuals in mundane situations (e.g., classroom settings) are less likely to report mixed emotions of happiness and sadness with dichotomous-then-unipolar scales than with simple unipolar scales (Russell & Carroll, 1999), thus making dichotomous-then-unipolar scales particularly useful for strong tests of whether people can experience mixed emotions in more emotionally complex situations. Indeed, Larsen et al. (2004) used dichotomous-then-unipolar scales in their study demonstrating that disappointing wins and relieving losses elicit ambivalence (see also, Larsen, McGraw, & Cacioppo, 2001). On the other hand, dichotomous-then-unipolar scales compound the problem of simple unipolar scales' inefficiency (Schimmack, 2005). Whereas simple unipolar scales require two responses to measure positivity and negativity, dichotomous-then-unipolar scales require at least two responses and as many as four. To report ambivalence, for example, respondents must endorse feeling both positive and negative and rate the intensity of that positivity and that negativity.

The evaluative space grid

We introduce the *evaluative space grid* (ESG; see Figure 1) as a single-item measure of positivity and negativity. The ESG is structurally similar to Russell, Weiss, and Mendelsohn's (1989) *affect grid*, a 10×10 grid in which respondents indicate how pleasant or unpleasant they feel along the x-axis and how aroused they feel along the y-axis. In the ESG, however, respondents indicate how positive and negative they feel along the x- and

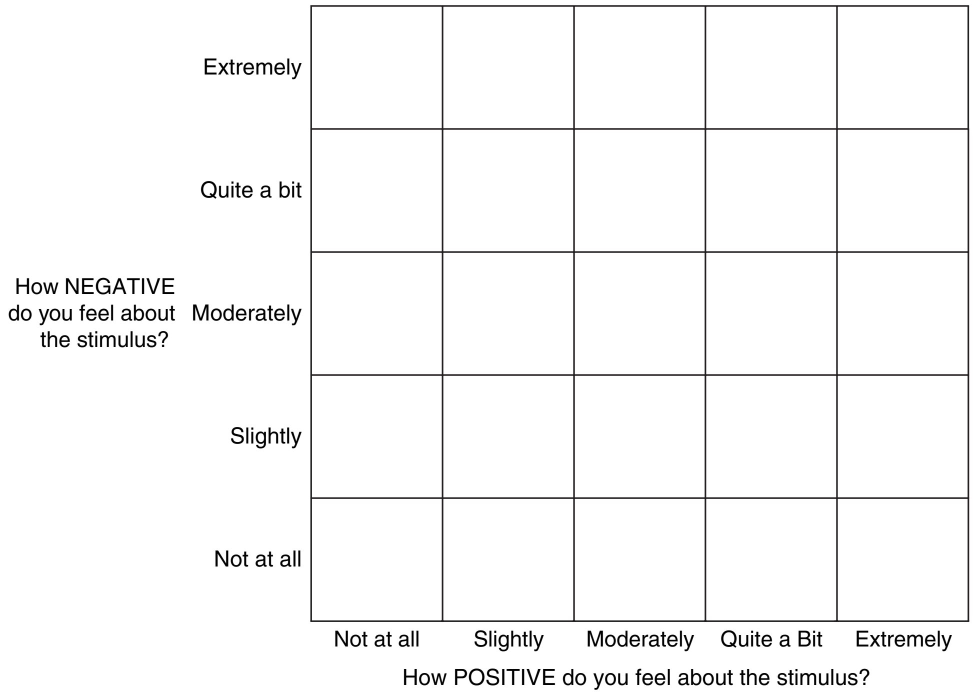

Figure 1. The evaluative space grid (ESG). After receiving instructions on how to use the ESG, the participant is presented with stimuli and rates each one in turn. The participant uses the mouse to select one of the grid's 25 cells. The positivity is taken as the number of the cell selected, with cells numbered along the *x*-axis from 0 to 4. The negativity score is taken as the number of the cell selected, with cells numbered along the *y*-axis from 0 to 4.

y-axes. Respondents can be provided with any number of levels of positivity and negativity. In all studies reported here, the grid was computerised and respondents' task was to click on the cell that best captured their evaluative reactions (see the appendix for instructions). If the grid is presented on paper, respondents can make a mark on the appropriate cell.

We report four studies assessing the ESG's convergent validity, predictive validity, and generalisability. Specifically, we compared ratings from the grid with ratings from currently available unipolar measures of positivity and negativity (Studies 1 & 4), bipolar measures of valence (Studies 3 & 4), and facial electromyographic markers of positive and negative affect (Study 4). In addition, in Study 2 we investigated the ESG's ability to track the continuous flux of positivity and negativity over time.

STUDY 1

To examine the grid's convergent validity, in Study 1 we compared ESG ratings of gamble outcomes with ratings obtained from simple unipolar and dichotomous-then-unipolar scales. We also examined whether participants could complete the grid more quickly than they could complete simple unipolar and dichotomous-then-unipolar scales.

Method

Participants. Participants were 62 Texas Tech University undergraduates who completed the study in exchange for course credit. One participant reported feeling good about only one of 32 gamble outcomes and never reported feeling bad; another participant appeared to respond randomly. Their data were removed, leaving a sample size of 60 (26 women; 43%). Of these, 18, 21, and 21 participants had been randomly assigned to the ESG, simple unipolar scale, and dichotomous-then-unipolar scale conditions, respectively

Stimuli. Participants played 32 computerised 50–50 gambles. There were eight disappointing wins, in which participants won the smaller of two amounts. Specifically, they won \$2, \$3, \$3, \$5, \$5, \$8, \$8, and \$12 instead of \$5, \$5, \$8, \$8, \$12, \$12, and \$17, respectively. There were also eight outright wins, in which participants won the larger of two amounts. Specifically, they won \$5, \$5, \$8, \$8, \$12, \$12, \$17, and \$17 instead of \$2, \$3, \$3, \$5, \$5, \$8, \$8, and \$12, respectively. Losing gambles were constructed by reversing the signs of the outcomes. This yielded eight relieving losses, in which participants lost the smaller of two amounts, and eight outright losses, in which they lost the larger of two amounts.

Procedure. Participants were told they would rate how good and bad they felt about a series of card games. Participants in the ESG condition were given the instructions in the appendix and rated their positive and negative reactions on a 5×5 grid. The left (top), middle (middle), and bottom (right) cells were labelled "*not at all*", "*moderately*", and "*extremely*", respectively. Participants in the simple unipolar condition were instructed to rate how good and how bad they felt about each outcome on scales ranging from 0 to 4. The left, middle, and right cells were labelled "*not at all*", "*moderately*", and "*extremely*", respectively. Participants in the dichotomous-then-unipolar condition were instructed to indicate whether they felt good or bad when prompted and, if so, how good or bad they felt on a scale from 1 to 4; the left and right cells were labelled "*slightly*" and "*extremely*", respectively. (An initial response of "no" was assigned a score of 0.) Half of the participants in the simple unipolar and dichotomous-then-unipolar conditions rated whether they felt good first and the remainder rated whether they felt bad first. After several practice card games, participants were given \$5 cash and told that the sum of their wins and losses would determine whether they won additional money or lost some or all of this endowment.

The order of the 32 gambles was randomised for each participant. At the beginning of each game, the stakes (e.g., win $5, win $8) appeared at the top of the screen, accompanied by two cards dealt facedown. After 3 s, the cards turned over, thereby revealing their values. The card containing the obtained outcome was designated with a red border. For instance, if the card with a red border read "win $5", the participant won $5. After another 3 s, participants made their ratings. Prompts appeared immediately after participants' previous responses in the simple unipolar and dichotomous-then-unipolar conditions. This allowed us to compare completion times in these conditions with those in the ESG condition, where participants made only one response.

Results

To assess the grid's convergent validity, we first compared ratings of outright wins, disappointing wins, relieving losses, and outright losses in the three response format conditions. As shown in Figure 2, ratings in the grid condition were similar to those in the simple unipolar and dichotomous-then-unipolar conditions. That is, whereas outright wins were generally rated positive and outright losses negative, disappointing wins and relieving losses were rated both positive and negative. To examine these patterns more systematically, we submitted positive and negative ratings to a 3 (Response Format: grid, simple unipolar, dichotomous-then-unipolar) × 4 (Outcome: outright win, disappointing win, relieving loss, outright loss) × 2 (Rating: positive, negative) mixed-model ANOVA in which response format was a between-subjects variable. This ANOVA revealed main effects of Outcome and Rating and an Outcome × Rating interaction. More important, however, there was no Response Format × Outcome × Rating interaction, $F(6, 112) = 1.40$, *ns*, indicating that the pattern of positive and negative ratings in the grid condition was comparable to those in the other two response format conditions.[1]

We conducted three sets of Bonferroni-corrected planned comparisons to decompose the Outcome × Rating interaction and to further compare patterns of ratings in the three response format conditions. Not surprisingly, outright wins were rated more positive and less negative than outright losses in all three conditions (see Figure 2). More important, both disappointing

[1] Unless otherwise noted, all reported effects are significant at $p < .05$, two-tailed, and all effects involving independent variables with 3 or more levels were evaluated with the Huyn–Feldt correction. The ANOVA also revealed a main effect of response format. Post hoc tests indicated that grid ratings ($M = 1.63$, $SD = 0.31$) were higher than dichotomous-then-unipolar ratings ($M = 1.23$, $SD = 0.15$) but lower than simple unipolar ratings ($M = 1.70$, $SD = 0.31$).

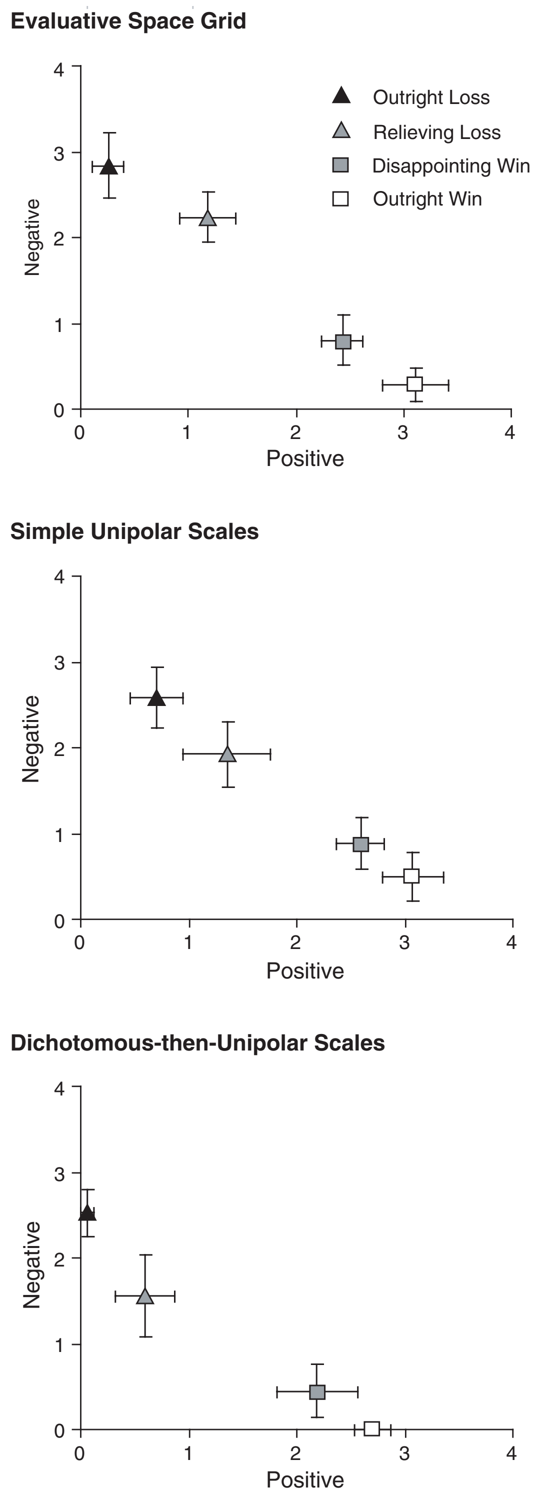

Figure 2. Mean positive and negative ratings of outright wins, disappointing wins, relieving losses, and outright losses in Study 1's ESG (top panel), simple unipolar (middle panel), and dichotomous-then-unipolar (bottom panel) conditions. Error bars represent 1.96 standard errors.

wins and relieving losses were rated more positive than outright losses in all three conditions. Moreover, both disappointing wins and relieving losses were rated more negative than outright wins. Whereas the results from the dichotomous-then-unipolar condition replicate those of Larsen et al. (2004), the generality of these patterns to the grid condition provides initial evidence for the ESG's convergent validity.

One limitation of simple unipolar scales is that people can misinterpret them and, as a result, report feeling positive even when there is little cause for positivity and negative even when there is little cause for negativity (see Russell & Carroll, 1999). Indeed, outright wins are unequivocally pleasant events, but participants in the simple unipolar condition often reported feeling negative about them ($M = 0.49$, $SD = 0.66$). Similarly, outright losses are unequivocally unpleasant, but participants in the simple unipolar condition often reported feeling positive ($M = 0.57$, $SD = 0.56$). To the extent that the ESG provides valid measures, negative ratings of outright wins and positive ratings of outright losses should be lower in the ESG condition than in the simple unipolar condition. To examine this hypothesis, we averaged the negative ratings of outright wins along with the positive ratings of outright losses into a single index of anomalous ratings. We submitted this variable to a one-way ANOVA with Condition as a between-subjects factor. The effect of Condition was significant and a pair of Dunnett post hoc comparisons indicated that participants in the simple unipolar condition ($M = 0.52$, $SD = 0.50$) gave higher anomalous ratings than did those in the ESG condition ($M = 0.25$, $SD = 0.29$), whose ratings were marginally higher than those in the dichotomous-then-unipolar condition ($M = 0.03$, $SD = 0.06$), $p = .09$. Thus, participants in the grid condition were less likely than those in the simple unipolar condition to rate outright wins as negative and outright losses as positive.

To further investigate the ESG's convergent validity, we examined correlations between ratings from the grid condition and those from the simple unipolar and dichotomous-then-unipolar conditions. Specifically, for each condition we computed mean ratings for all 32 gamble outcomes. As shown in Figure 3, positive and negative ratings from the grid were highly correlated with those from the simple unipolar (left panels) and dichotomous-then-unipolar response formats (right panels). Indeed, all correlations exceeded $r(32) = .97$, thereby providing further evidence for the ESG's convergent validity.

Completion times. To examine whether participants in the ESG condition made their ratings more quickly than those in the other response format conditions, we submitted mean log-transformed completion times,

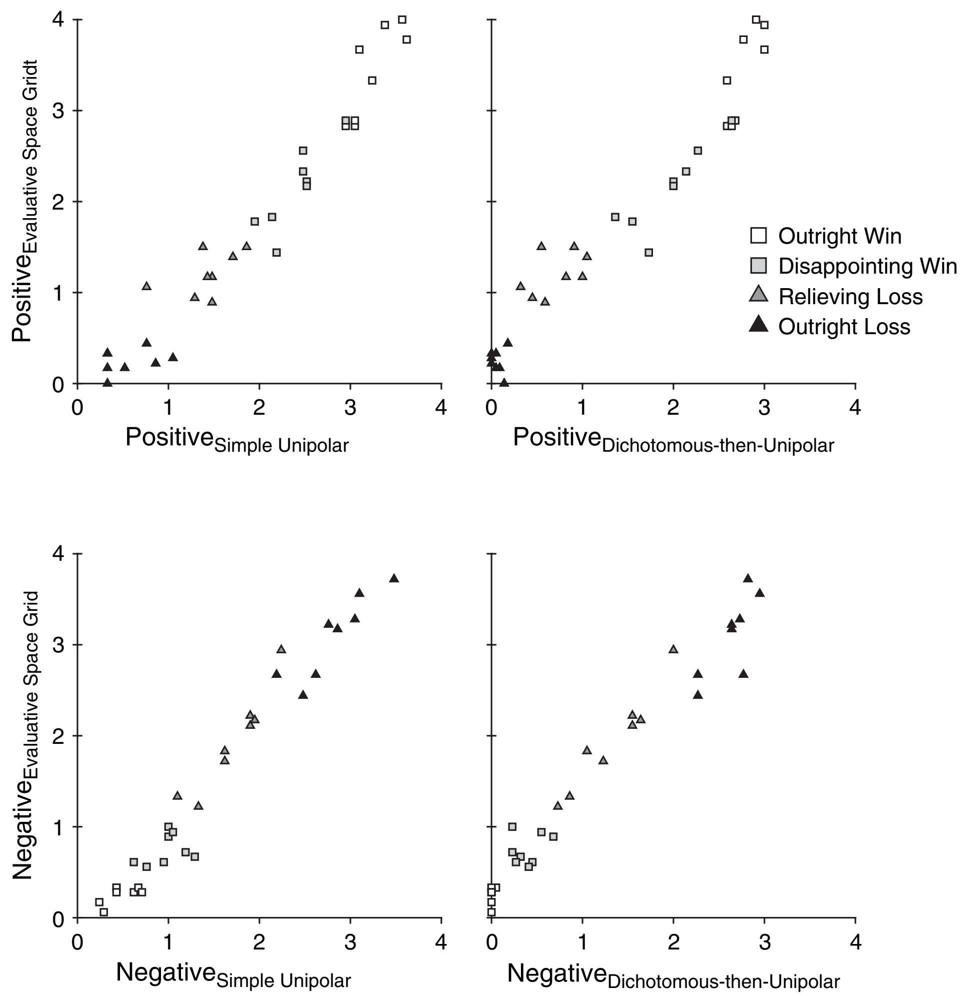

Figure 3. Mean positive ratings (top panels) and negative ratings (bottom panels) from the ESG as a function of mean simple unipolar ratings (left panels) and mean dichotomous-then-unipolar ratings (right panels) in Study 1.

collapsed across all 32 outcomes, to a one-way ANOVA with Response Format as the independent variable. In the simple unipolar and dichotomous-then-unipolar conditions, completion time was defined as the amount of time between the onset of the first prompt and the participant's response to the final prompt. The effect of Condition was significant and post hoc Dunnett comparisons revealed that completion times were lower in the grid condition (untransformed $M = 4.1$ s, $SD = 1.7$) than in the simple unipolar condition ($M = 5.3$ s, $SD = 1.7$; $d = 0.69$) and dichotomous-then-unipolar condition ($M = 7.3$ s, $SD = 2.9$; $d = 1.30$). Thus, the ESG provides a more efficient means of measuring positivity and negativity than do simple unipolar and dichotomous-then-unipolar scales.

Discussion

Study 1 provides initial evidence that the ESG provides valid measures of positivity and negativity. Along with participants in the simple unipolar and dichotomous-then-unipolar conditions, participants in the grid condition rated disappointing wins and relieving losses as eliciting both positivity and negativity. Moreover, mean positive and negative ratings in the ESG condition correlated almost perfectly with those in the other conditions.

Simple unipolar scales have been criticised because people often report ambivalent about non-ambivalent situations with them (Russell & Carroll, 1999). In Study 1, participants in the ESG condition rated unambiguously bad events as less positive and unambiguously good events as less negative than did those in the simple unipolar condition. Such anomalous ratings were, however, marginally higher in the ESG condition than in the dichotomous-then-unipolar condition. Thus, grid ratings appear to be more strictly unipolar than simple unipolar ratings, but may not be as strictly unipolar as dichotomous-then-unipolar ratings.

Study 1 also indicated that the ESG has at least one advantage over both simple unipolar and dichotomous-then-unipolar scales. Specifically, it took 27% and 47% less time to complete the ESG than it took to complete simple unipolar and dichotomous-then-unipolar scales, respectively. In studies with large numbers of stimuli, this difference can be substantial. In a study with 200 stimuli, for instance, the use of the ESG, as opposed to dichotomous-then-unipolar scales, would trim > 10 min from the study, which may minimise participant fatigue and boredom.

It was unfeasible in Study 1 to compare the ESG with all of the many unipolar measures currently available, such as Schimmack's (2001, 2005) modified simple unipolar scales. Schimmack presented participants with simple unipolar scales but took steps to reduce the likelihood that participants would mistake them for bipolar scales. For instance, he explicitly instructed participants to consider whether they felt each emotion before considering its intensity. The extent to which the ESG and Schimmack's modified simple unipolar scales yield comparable data remains unclear, but Study 1's finding that participants can complete the grid more quickly than simple unipolar scales suggests that they would also complete the grid more quickly than modified simple unipolar scales.

STUDY 2

In Study 2, we extended Study 1's finding that participants can complete the ESG more quickly than other measures by examining whether the ESG can provide continuous assessments of positivity and negativity over time. Both dichotomous-then-unipolar and simple unipolar scales may offer limited temporal resolution because they can take a fairly long time to complete. Even if instructed to indicate how they feel at this very moment, participants might aggregate their evaluative reactions over some amount of time and subsequently report ambivalence even if, in fact, they had merely vacillated between positivity and negativity. Larsen et al. (2004) overcame this limitation in one experiment by asking participants to press one button whenever they felt good about how each card game was turning out and another button whenever they felt bad. The state of each mouse button was recorded every 100 ms, thereby providing continuous measures of positivity and negativity. Participants spent more time pressing both buttons in response to disappointing wins and relieving losses as opposed to outright wins and relieving losses, thus demonstrating that mixed outcomes simultaneously elicit positivity and negativity.

Whereas Larsen et al.'s (2004) button-press measures only provide dichotomous continuous measures of positivity and negativity, in Study 2 we examined whether the ESG could be adapted to provide measures that were both graded and continuous. Specifically, we asked participants to report their moment-by-moment evaluative reactions to computerised card games by moving the mouse throughout the ESG while the computer recorded the mouse location every 100 ms. Based on Larsen et al.'s evidence that disappointing wins and relieving losses simultaneously elicit positivity and negativity, we expected participants to report more ambivalence in response to disappointing wins and relieving losses as opposed to outright wins and losses. That is, we expected them to move the mouse farther into the interior of the *continuous evaluative space grid* in response to disappointing wins and relieving losses.

Method

Participants. Participants were 19 Ohio State University undergraduates who completed the study in exchange for course credit.

Stimuli. Participants played 16 50–50 gambles identical to those of Larsen et al. (2004) in one of two random orders. Disappointing wins were wins of $5 instead of $6, $9, or $12. In addition, there was a single outright win of $5 instead of $3. Losing gambles were constructed by reversing the signs of the outcomes. This resulted in relieving losses of $5 instead of $6, $9, or $12 and a single outright loss of $5 instead of $3. So that participants would not be suspicious as to why the obtained outcome was always a win or loss of $5, we also included filler gambles by swapping the eight target gambles' obtained and unobtained outcomes. In light of Larsen et al.'s (2004) finding that mixed outcomes only began to elicit ambivalence several seconds after the resolution of the gambles, we extended the duration of the payoff period from 3 s to 6 s.

Procedure. Participants were given instructions similar to those in the appendix with the exception that they were asked to indicate their momentary positive and negative reactions with a 7×7 grid as the card games unfolded over time. Participants were informed that the computer would record the mouse location 10 times every second (i.e., every 100 ms), thereby allowing them to move the mouse as quickly as they wished. Positive and negative ratings were aggregated across the 60 samples collected during the payoff period.

Results

Figure 4 shows aggregate positive and negative ratings from the 6 s payoff period. Data are collapsed across the three disappointing wins and relieving losses. These data were submitted to a 2 (Trial Order: A, B) $\times$ 2 (Outcome: outright loss, relieving loss, disappointing win, outright win) $\times$ 2 (Rating: positive, negative) mixed-model ANOVA, where Trial Order was manipulated between-subjects. The ANOVA revealed main effects of Outcome and Rating, both of which were qualified by an Outcome $\times$ Rating interaction. Bonferroni-corrected planned comparisons indicated that outright wins were rated more positive and less negative than outright losses, $d = 3.17$ and 2.07, respectively (see Figure 4). More important, though relieving losses were not

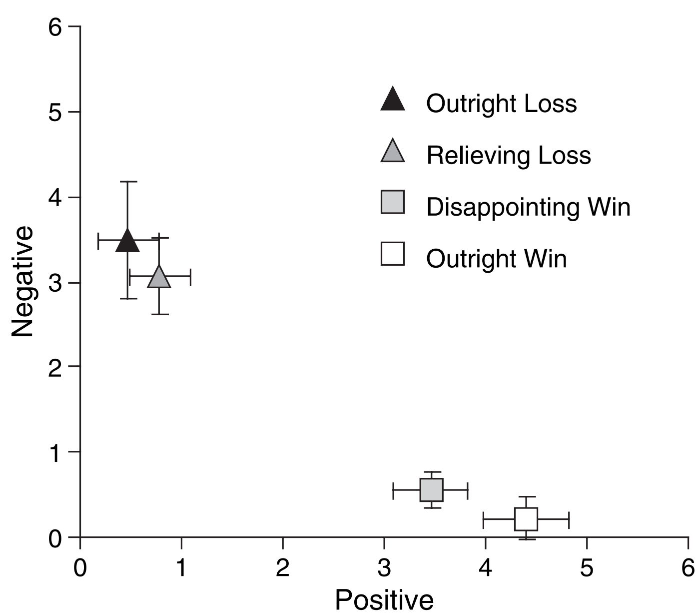

Figure 4. Mean positive and negative ratings of outright wins, disappointing wins, relieving losses, and outright losses during the 6 s payoff period in Study 2. Error bars represent 1.96 standard errors.

rated significantly more positive than outright losses, $d = 0.52$ (uncorrected $p = .03$), disappointing wins were rated more positive than outright losses, $d = 3.02$. In addition, both disappointing wins and relieving losses were rated more negative than outright wins, $d = 2.77$ and 0.74, respectively.

To examine whether participants experienced more ambivalence in response to disappointing wins and relieving losses, we quantified how far they moved into the interior of the grid. Specifically, we computed the minimum of each participant's positive and negative ratings during each sample of each card game, i.e., MIN (P, N). MIN scores provide a graded index of ambivalence by taking on values of 0 when participants rate stimuli as neutral, exclusively positive, or exclusively negative, but higher values when participants rate stimuli as both positive and negative (Kaplan, 1972; Schimmack, 2001). Mean MIN scores, which are shown in Table 1, were submitted to a 2 (Trial Order: A, B) $\times$ 2 (Outcome: outright loss, relieving loss, disappointing win, outright win) mixed-model ANOVA. The ANOVA revealed a main effect of Outcome and a planned comparison revealed that MIN scores were greater during disappointing wins and relieving losses ($M = 0.58$, $SD = 0.43$) than during outright wins and losses ($M = 0.28$, $SD = 0.45$), $d = 0.85$.

One possibility is that participants made only brief excursions into the interior of the ESG in the course of reporting vacillating feelings of positivity and negativity. If so, one would expect instances of mixed feelings to be numerous and short-lived, but there were never more than two excursions into the interior during any one card game and the average duration of each excursion was 3.4 s ($SD = 1.5$).

Discussion

Study 2's results demonstrated that the continuous ESG provides measures of positivity and negativity that are both graded (like simple unipolar and dichotomous-then-unipolar scales) and continuous (like Larsen et al.'s, 2004, button-press measures). Indeed, just as Larsen et al. found that people often pressed the good and bad buttons simultaneously in response to mixed

TABLE 1
MIN scores in response to outright losses, relieving losses, disappointing wins, and outright wins (Study 2)

Outcome	*Mean*	*SD*
Outright loss	0.33	0.56
Relieving loss	0.60	0.55
Disappointing win	0.56	0.47
Outright win	0.22	0.54

outcomes, Study 2's participants often moved the cursor into the interior of the grid in response to mixed outcomes.

STUDY 3

In Study 3, we investigated the relationship between ESG ratings and bipolar valence ratings. Bipolar valence is a function of the net difference between positivity and negativity (e.g., Cacioppo & Berntson, 1994). Thus, the net difference between the ESG's positive and negative ratings (i.e., P – N) should be highly correlated with bipolar valence ratings. Such findings would provide evidence for the grid's predictive validity.

We also investigated a potential advantage of the ESG over bipolar valence ratings. Middling ratings on bipolar scales are ambiguous because they can reflect either neutrality (i.e., the absence of both positivity and negativity) or ambivalence. By allowing individuals to report their positive and negative reactions separately, the ESG should be able to disambiguate neutrality from ambivalence. In Studies 1 and 2 we used gamble outcomes as stimuli, but the grid should generalise to other evaluative stimuli, including attitude objects. To examine these three possibilities, we asked participants to complete the ESG and Russell et al.'s (1989) affect grid in response to a variety of attitude objects.

Method

Participants. Participants were 18 (12 female; 67%) University of Chicago undergraduates who were paid for their time. Participants were recruited through e-mail and campus advertisements. Four additional participants' data were lost due to computer error and another 12 individuals completed the preliminary questionnaire but failed to complete the experiment.

Procedure. Individuals who responded to the advertisements were e-mailed a questionnaire containing seven lists (e.g., "List 10 people or things you care about"). The sixth list asked people to list 10 things that they felt ambivalent about. *Bill Clinton*, *capital punishment*, and *exercise* were provided as examples. Participants e-mailed their completed lists back to the experimenter, who then scheduled an experimental session between 6 and 10 days later and selected eight of the ten ambivalent objects to include as stimuli in that participant's stimulus set. All other lists were ignored. In addition to the eight ideographically selected ambivalent attitude objects, we devised three lists of eight attitude objects assumed to be negative (e.g., *terrorism*, *traffic jams*), neutral (e.g., *lettuce*, *wallpaper*), and positive (e.g., *spring break*, *sunshine*) for the majority of participants. Upon arrival, participants were randomly assigned to complete the ESG task before ($n = 8$) or after ($n = 10$) the affect grid task. The ESG and affect grids were 5×5 and 10×10, respectively. Before each task, participants received instructions on how to use the measure and then rated all 32 attitude objects in a single random order.[2]

Results

To assess bipolar evaluative reactions to the attitude objects, we collapsed valence ratings from the affect grid across the eight exemplars in each valence category. These mean ratings were submitted to a 2 (Order of ESG Task: first, second) $\times$ 4 (Valence Category: negative, neutral, positive, ambivalent) mixed-model ANOVA in which Order was a between-subjects variable. The ANOVA revealed a main effect of Valence Category (see Figure 5), which we decomposed by conducting all possible Bonferroni-corrected contrasts. Not surprisingly, positive objects ($M = 2.74$, $SD = 0.70$) received higher valence ratings than negative objects ($M = -2.77$, $SD = 0.54$), $d = 4.65$. More important, both neutral objects ($M = 0.67$, $SD = 0.48$) and ambivalent objects ($M = -0.06$, $SD = 0.68$) received higher valence ratings than negative objects, ds = 4.25 and 3.49, respectively, but lower ratings than positive objects, ds = 2.79 and 2.87, respectively. In addition, neutral objects received higher ratings than ambivalent objects, $d = 1.02$. Thus, bipolar valence ratings suggest that reactions to ambivalent objects came closer to the neutral point than did reactions to neutral objects.[3]

Differences in bipolar valence, however, failed to capture the defining difference between neutral and ambivalent stimuli: whereas neutral stimuli elicited little positivity or negativity, ambivalent stimuli elicited both positivity and negativity. Thus, ESG data should reveal that ambivalent objects are rated both more positive and negative than neutral objects. Mean positive and negative ratings collapsed across the eight exemplars in each valence category are shown in the bottom panel of Figure 5. These mean ratings were submitted to a 2 (Order of ESG Task: first, second) $\times$ 4 (Valence Category: negative, neutral, positive, ambivalent) $\times$ 2 (Rating: negative,

[2] Participants also completed a gambles task similar to that of Studies 1 and 2. Results from the gambles task are excluded due to space constraints.

[3] An analogous ANOVA on the affect grid's arousal ratings yielded a main effect of valence category. Bonferroni-corrected post hoc comparisons revealed that negative, ambivalent, and positive attitude objects were rated more arousing than neutral attitude objects. No other contrasts were significant. The finding that negative and positive attitude objects were rated more arousing than neutral attitude objects replicates the curvilinear effect of valence on arousal (Lang, Greenwald, Bradley, & Hamm, 1993). The finding that ambivalent attitude objects were rated more arousing than neutral attitude objects is more novel, but is consistent with the hypothesis that ambivalence is uncomfortable (Cacioppo & Berntson, 1994).

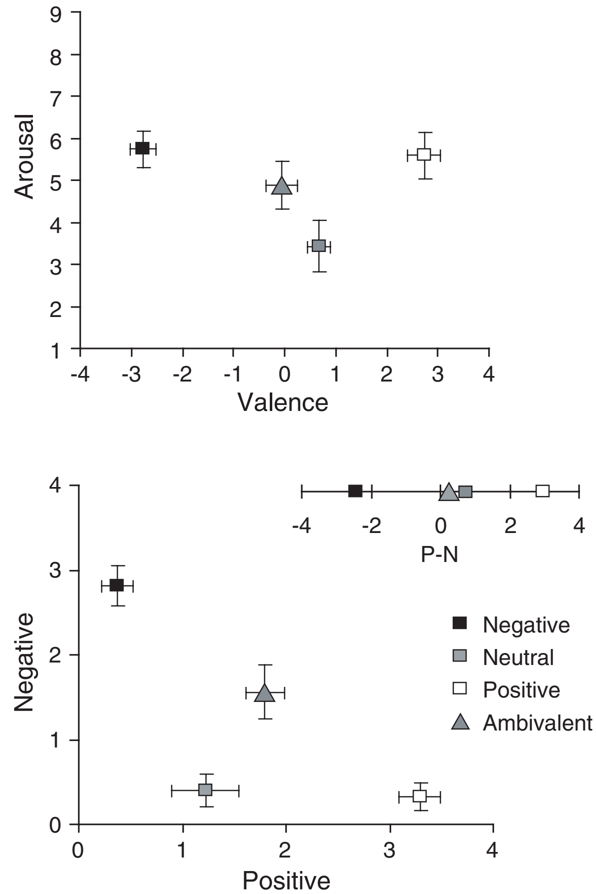

Figure 5. Mean valence and arousal ratings from the affect grid (top panel) and positive and negative ratings from the ESG (bottom panel) for negative, neutral, positive, and ambivalent attitude objects in Study 3. The inset displays the net difference between mean positive and negative ratings for the four types of attitude objects. Error bars represent 1.96 standard errors.

positive) mixed-model ANOVA. The ANOVA revealed main effects of Valence Category and Rating, both of which were qualified by a Valence Category × Rating interaction. We conducted six Bonferroni-corrected planned comparisons to decompose the interaction. Not surprisingly, positive objects were rated more positive ($M = 3.30$, $SD = 0.43$) and less negative ($M = 0.33$, $SD = 0.37$) than negative objects ($M_{\text{Positivity}} = 0.38$, $SD = 0.33$; $M_{\text{Negativity}} = 2.82$, $SD = 0.51$), $d = 5.75$ and -4.10, respectively. Neutral objects were rated no more negative ($M = 0.40$, $SD = 0.42$) than positive objects, $d = 0.17$, but they were unexpectedly rated more positive ($M = 1.22$, $SD = 0.70$) than negative objects, $d = 1.21$. Moreover, ambivalent objects were rated both more positive ($M = 1.80$, $SD = 0.40$) and more negative ($M = 1.56$, $SD = 0.69$) than neutral objects, $d = 0.74$ and 1.43, respectively. The latter finding is most important because it demonstrates that the ESG, unlike the affect grid, can capture the differences in positive and negative reactions between neutral and ambivalent objects.

To examine the ESG's predictive validity, we compared bipolar valence ratings from the affect grid with ESG ratings. Figure 5's inset shows that the rank order of the affect grid's mean valence ratings for the four valence categories can be reconstructed by subtracting mean ESG negative ratings from mean ESG positive ratings. In addition, valence and P – N ratings were highly correlated for the average participant ($M = 0.90$, $SD = 0.05$). Collapsing ratings across participants is complicated by the fact that we selected ambivalent objects ideographically, but we computed proxy mean ratings for those stimuli by sorting each participant's ambivalent objects by valence rating. In any event, the correlation between mean valence and mean P – N ratings approached unity whether proxy ambivalent attitude objects were included or excluded, both $r\text{s}(32) = .99$.

Discussion

As evidenced by Study 3's finding that participants rated ambivalent attitude objects as both more positive and negative than neutral attitude objects, the ESG can capture the defining difference between ambivalent and neutral attitudes. In addition, the finding that the net difference between the grid's positive and negative ratings is almost perfectly correlated with the affect grid's valence ratings provides evidence for the ESG's generalisability to attitudes and its predictive validity. It also indicates that a single ESG rating conveys all the information that a bipolar valence rating does, as well as information about the underlying levels of positivity and negativity.

STUDY 4

In Study 4, we focused on the grid's ability to assess affective reactions. The tripartite model of attitudes (e.g., Breckler, 1984) holds that cognitive reactions (i.e., beliefs), affective reactions (i.e., emotions), and behavioural reactions (e.g., approach vs. avoidance) provide the bases for evaluative reactions. Gamble outcomes like those used in Studies 1 and 2 can elicit emotions (e.g., Mellers et al., 1997), but evaluations of gamble outcomes presumably also have a substantial cognitive basis in that money represents an abstract, secondary reinforcer. Study 3's attitude objects might also have elicited predominantly cognitive reactions because, for instance, participants were presented with the objects' verbal labels, rather than the objects themselves (Breckler, 1984). Relative to words (e.g., *murder*), sights and sounds (e.g., witnessing a murder) can elicit a great deal of affect. For example, unpleasant pictures and sounds potentiate the startle eye blink response (e.g., Lang, Bradley, & Cuthbert, 1990), but unpleasant words do not appear to do so (Aitken, Siddle, & Lipp, 1999). Thus, in Study 4 we asked participants to rate pictures and sounds drawn from normed and standardised stimulus sets. For comparison purposes, we also asked them to rate valenced words. To the extent that the grid's validity generalises to affectively charged stimuli, correlations between grid ratings and normative ratings should be as high in the pictures and sounds tasks as they are in the words task.

We also measured facial electromyographic (EMG) activity over *zygomaticus major* and *corrugator supercilii*, which is associated with evaluative reactions and is thought to be most strongly associated with affectively based evaluative reactions, in particular (Cacioppo, Petty, Losche, & Kim, 1986; Lang, Greenwald, Bradley, & Hamm, 1993). Whereas *zygomaticus major* pulls the corners of the mouth back and up into a smile, *corrugator supercilii* pulls the brow into a frown. Increased activity over the cheek and brow provide markers for positive and negative affect, respectively (Cacioppo et al., 1986; Lang et al., 1993). In addition, decreased activity over the brow provides a marker for positive affect (Lang et al., 1993; Larsen, Norris, & Cacioppo, 2003). To the extent that the ESG can index evaluative reactions to affectively charged stimuli, ESG ratings of pictures and sounds should be associated with EMG activity. In that evaluations of words have less of an affective basis, ESG ratings of words should be less strongly associated with EMG activity.

Evaluations of valenced pictures are stable over time (Stark et al., 2004). Thus, grid ratings of pictures, and perhaps sounds and words, should also be stable. To examine the temporal stability of ESG ratings, we asked participants to rate each stimulus twice over the course of two weeks.

Method

Participants. Sixty-eight University of Chicago undergraduate women participated for payment. Only testing women allowed us to use the same set of stimuli for all participants, as the project's primary aim was to study individual differences in evaluative processing.[4] Participants were fluent in English, healthy, and not currently taking any psychotropic medication. Data from three participants were removed because they failed to complete both sessions. In addition, six data files from one or the other session were lost due to equipment malfunction or experimenter error. In these cases, data from the relevant task were removed but data from the other two tasks were

[4] Larsen et al. (2003) also reported results from this study, but did not examine aspects of the data relevant to the results reported here (e.g., temporal stability of evaluative space grid ratings, task effects on EMG activity).

retained. The final sample contained 63, 62, and 64 participants for the pictures, sounds, and words tasks, respectively.

Stimulus materials. Sixty-six colour pictures, sound clips, and words were selected from the International Affective Picture System (IAPS; Center for the Study of Emotion and Attention, 1999; Lang, Bradley, & Cuthbert, 1999), International Affective Digitised Sounds (IADS; Bradley & Lang, 1999b), and Affective Norms for English Words (ANEW; Bradley & Lang, 1999a), respectively, based on each stimulus set's normative valence ratings. Stimuli spanned the bipolar valence dimension and varied widely in content.

Procedure. Following an orientation session, participants completed two experimental sessions separated by about two weeks ($M = 15.4$ days). In each experimental session, all participants completed the pictures, sounds, and words tasks, and an additional task to be reported elsewhere. The tasks were presented in the following random orders: words, pictures, sounds (Session 1), and sounds, pictures, words (Session 2). Stimuli were presented in the same random order in Session 1 and in another random order in Session 2. Trials consisted of a 3 s baseline period, a 6 s stimulus period, and a 3 s recovery period. Following the recovery period, participants used a 5×5 ESG to rate their positive and negative reactions. Facial EMG was recorded over the right cheek and brow with 4 mm standard Ag/AgCl electrodes, amplified $5000\times$, and digitised at 1000 Hz.

EMG data reduction. Offline, data were submitted to a 15 Hz high-pass filter and full-rectified, then visually inspected for artefact. To correct for the positive skew inherent to EMG data, data were then subjected to a square-root transformation. Following Lang et al. (1993), EMG reactivity was measured as the difference between activity during the 6 s stimulus period and the 1 s preceding stimulus onset.

Results

To examine the stability of ESG ratings, for each participant we computed correlations between their ratings of each stimulus in Sessions 1 and 2. As shown in Table 2, mean correlations were strong for positive and negative ratings in all three tasks. Thus, ESG ratings are stable over time.

To examine how well the ESG generalises to affectively charged stimuli (e.g., pictures, sounds), we first conducted analyses similar to those reported in Study 3. Specifically, we measured how well the difference between mean positive and negative ratings (i.e., P – N), collapsed across sessions, was correlated with normative valence ratings in the pictures and sounds task as

TABLE 2
Mean correlations between ratings in Sessions 1 and 2 (Study 4)

	Rating			
	Positive		*Negative*	
Task	*Mean*	*SD*	*Mean*	*SD*
Pictures	0.80	0.10	0.83	0.10
Sounds	0.73	0.10	0.73	0.10
Words	0.74	0.12	0.75	0.13

opposed to the words task. Normative valence ratings of the pictures, sounds, and words were obtained from Lang et al. (1999), Bradley and Lang (1999a), and Bradley and Lang (1999b), respectively. Replicating the findings from Study 3's attitudes task, in the words task the correlation between P – N and normative valence ratings approached unity, $r(66) = .96$. Moreover, the correlation between P – N and normative valence ratings was just as strong in the pictures task, $r(66) = .94$, and sounds tasks, $r(66) = .93$. Ito et al. (1998) provided normative unipolar positivity and negativity ratings for the majority of pictures presented in our pictures task. To investigate the ESG's convergent validity, we examined how well mean positive and negative ratings from the ESG were correlated with Ito et al.'s normative positive and negative ratings. Both correlations approached unity, $r(50) = .92$ and .90 for positive and negative ratings, respectively. Taken together, these findings provide evidence for the ESG's predictive validity for ratings of pictures and sounds and its convergent validity for ratings of pictures.

To address more directly the ESG's ability to measure affective reactions, we compared the strength of the relationships between participants' positive and negative ratings from the ESG and their EMG activity over the cheek and brow in the pictures, sounds, and words task. We used multilevel regression analysis to account for variance between participants (Level 2) while examining the Level 1 within-subject nested effect of task (pictures, sounds, words) on the relationship between ratings and EMG activity (Kreft & de Leeuw, 1998). Positivity and negativity ratings were simultaneously entered into the model to permit examination of the potentially independent effects of positivity and negativity on EMG activity.

Using EMG activity over the cheek as an example, the Level 1 equation is:

$$Y_{ij} = a_j + b_1 P_{ij} + b_2 N_{ij} + e_{ij},$$

where Y_{ij} is the grand mean level of EMG activity over the cheek across ratings (i) and participants (j), a_j is mean EMG activity for participant j, b_1 is a slope linking positivity ratings (P_{ij}) to EMG activity, b_2 is a slope linking negativity ratings (N_{ij}) to EMG activity, and e_{ij} is residual variance at Level 1. At Level 2, the equation for the intercept is:

$$a_j = \gamma_{00} + u_{0j},$$

where γ_{00} is the mean of the intercepts across subjects *j*, and u_{0j} represents deviations from the mean for participant *j*. The *t*-statistic corresponding to each parameter estimate is equivalent to the *z*-statistic and is provided by *B*/*SE*. Participants completed three qualitatively separate tasks so it was necessary to specify one of the tasks as the reference; we specified the words task as the reference in order to contrast the words task with the pictures and sounds tasks.

The multilevel model for activity over the brow revealed main effects of negativity and positivity, such that larger negative ratings were associated with greater activity over the brow and larger positive ratings were associated with less activity (see Figure 6). In addition, a main effect of task showed that pictures and sounds elicited greater activity over the brow than words. More important, the main effects of negativity and positivity were qualified by two-way interactions with task. As shown in the left panel of Figure 6, negativity potentiated activity over the brow more in the pictures task ($B = 0.025$, $SE = 0.003$) and sounds task ($B = 0.019$, $SE = 0.004$) than in the words task ($B = 0.008$, $SE = 0.002$). In addition, as shown in the right panel of Figure 6, positivity inhibited activity over the brow more in the pictures task ($B = -0.031$, $SE = 0.003$) and sounds task ($B = -0.048$; $SE = 0.004$) than in the words task ($B = -0.008$, $SE = 0.002$).

The effect of positivity on activity over the cheek is curvilinear such that only the most pleasant stimuli potentiate activity (Lang et al., 1993; Larsen

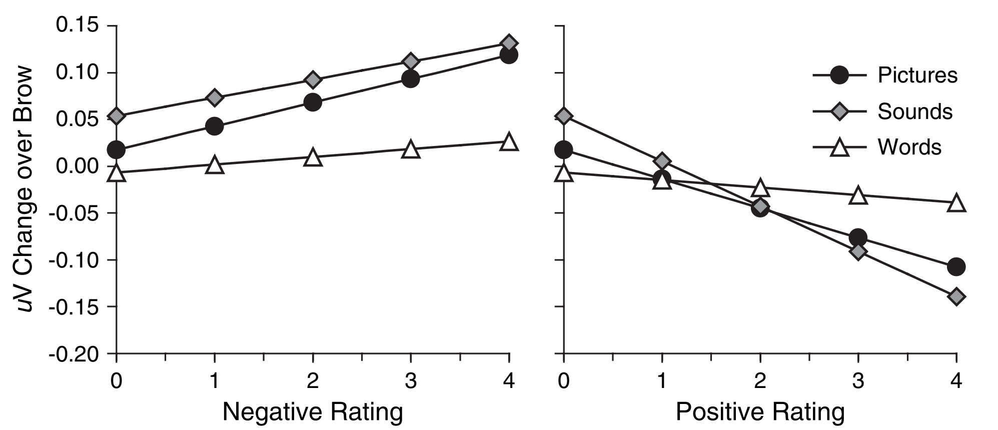

Figure 6. Predicted change in EMG activity over the brow as a function of negative ratings (left panel) and positive ratings (right panel) with separate curves for pictures, sounds, and words (Study 4). Positive and negative ratings are held constant at 0 in the left and right panels, respectively.

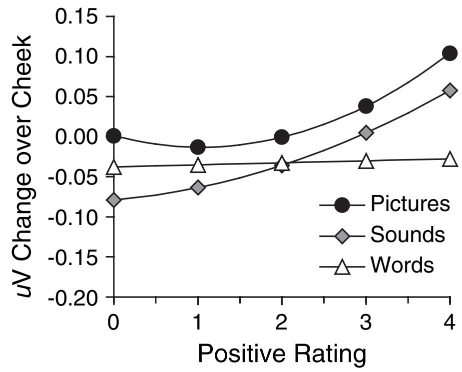

Figure 7. Predicted change in EMG activity over the cheek as a function of positive ratings with separate curves for pictures, sounds, and words (Study 4). Negative ratings are held constant at 0.

et al., 2003). To account for this curvilinearity, we included a quadratic term for positivity by adding squared grand-mean centred positivity ratings to the model (e.g., Raudenbush & Bryk, 2002). The analysis revealed significant linear and quadratic effects of positivity. As shown in Figure 7, greater positivity ratings were associated with increasingly more activity over the cheek. In addition, a main effect of task showed that pictures (but not sounds) elicited greater activity over the cheek than words. The linear and quadratic effects of positivity were qualified by two-way interactions with task. Indeed, in the words task neither the linear effect ($B = 0.003$, $SE = 0.003$) nor the quadratic effect ($B = 0.000$, $SE = 0.002$) approached significance. Moreover, the linear effect was stronger in the sounds task ($B = 0.006$, $SE = 0.003$) than in the words task and the quadratic effect was marginally stronger in the sounds task ($B = 0.006$, $SE = 0.003$) than in the words task ($p = .06$; see Figure 7). In addition, the quadratic effect was stronger in the pictures task ($B = 0.013$, $SE = 0.002$) than in the words task (see Figure 7). Neither the main effect of negativity nor the task × negativity interaction approached significance.

Discussion

Study 4 demonstrated that ESG ratings are stable and that the grid's generalisability extends to relatively affective stimuli (e.g., pictures, sounds). For example, the difference between mean positive and negative ESG ratings was just as highly correlated with normative valence ratings in the pictures and sounds tasks as in the words task. Moreover, positivity and negativity ratings showed expected patterns of relationships with facial EMG activity. Thus, the ESG captures affective reactions in particular as well as evaluative reactions in general.

GENERAL DISCUSSION

We conducted four studies involving a wide array of stimuli to examine the utility of the ESG as a single-item measure of positivity and negativity. Taken together, the studies provided evidence for the ESG's convergent and predictive validity as well as its efficiency and generalisability. For example, Studies 1 and 2 showed that ratings of gamble outcomes obtained with the ESG paralleled those obtained with more time-consuming unipolar measures. Moreover, Study 3 revealed that the ESG, unlike bipolar valence scales, can detect ambivalent attitudes and that valence ratings can be derived by simply taking the net difference between the ESG's positive and negative ratings. Among other findings, Study 4 demonstrated that ESG ratings are stable and predict psychophysiological markers of affect, particularly in tasks involving strongly affective stimuli.

With the continuous ESG, which was introduced in Study 2, positivity and negativity ratings can be collected many times each second. Such temporal resolution makes the continuous ESG especially useful for examining the simultaneous experience of mixed emotions. Following Larsen et al.'s (2001) demonstration that people often report feeling both happy and sad after *Life Is Beautiful*, for instance, Larsen and McGraw (2008) recently asked participants to complete the continuous ESG while watching clips from that film. Results indicated that participants moved farther into the interior grid while watching a clip that contained ostensibly bittersweet scenes than while watching a control clip, thereby providing additional evidence that people can feel happy and sad at the same time.

Even when continuous measures of positivity and negativity are not necessary, the standard ESG can be administered quickly, thereby rendering it useful for studies involving numerous stimuli (e.g., Study 4). Similarly, with a touch of a stylus the ESG could provide quick and minimally intrusive measures of positivity and negativity in experience sampling method studies. One caveat is that the ESG may be less useful as an efficient measure when only a few assessments are required, as the instructions for the ESG take approximately 90 s. Even in such situations, however, the ESG may provide less reactive and more valid measures if more time-consuming measures become so tedious that they actually alter evaluative reactions or reports.

In sum, the data indicate that the grid has a number of advantages over other measures of evaluative reactions. Unlike bipolar valence scales, it can detect ambivalence. In addition, participants can complete it more quickly than other unipolar scales. Finally, unlike other unipolar scales it can be used as a continuous measure of positivity and negativity.

Additional research with the ESG

Other labs have also begun using the grid. Hunter, Schellenberg, and Schimmack (2008) examined emotional reactions to music that varied in tempo and tone. Past research demonstrated that fast tempo elicits happiness and slow tempo elicits sadness. Similarly, the major tone elicits happiness and the minor tone elicits sadness. Hunter et al. investigated emotional reactions to music clips with conflicting cues (e.g., fast music in minor tones). Several studies revealed that music with conflicting cues elicited mixed emotions of happiness and sadness, as assessed both with simple unipolar scales and with the ESG. In another study, Andrade and Cohen (2007) used the continuous ESG to investigate emotional reactions to horror films. They found that horror films elicited more mixed emotions of happiness and fear among fans of the genre than among non-fans.

As would be expected if ambivalence is experienced infrequently (Cacioppo & Berntson, 1994), the ESG does not always reveal coactivation. Duncan, Barrett, and Russell (2006) had participants complete the continuous ESG as they simultaneously (a) listened to happy or sad music and (b) watched a series of pleasant and unpleasant pictures. Results indicated that participants rarely reported mixed emotions even while being exposed to music and pictures that were opposite in valence. Taken together, the findings of Hunter et al. (2008) Andrade and Cohen (2007), and Duncan et al. suggest that opposite-valence cues are more likely to elicit mixed emotions if they come from the same modality (e.g., dimensions of music; Hunter et al., 2008) than from different modalities (e.g., music vs. pictures; Duncan et al., 2006).

These researchers' studies indicate that the ESG has the potential to become a staple measure when investigators require measures of positivity and negativity in general or single pairs of opposite-valence emotions (e.g., happy, fear; Andrade & Cohen, 2007). In future emotion research, researchers might find it worthwhile to present several grids after each stimulus in order to gather ratings of several pairs of emotional reactions (e.g., happy vs. sad; calm vs. tense). Future research can also determine whether individuals other than undergraduates can use the grid effectively. Encouraging findings come from a recent study in which individuals who ranged in age from 30 to 80 years and who varied widely in education and socioeconomic status successfully used the grid to rate IAPS pictures (Norris, van Reekum, & Davidson, 2007).

Conclusion

The availability of unipolar measures of positive and negative evaluative reactions has fostered research on the separability of positivity and negativity, but further progress may require advances in measurement. To that end, we have shown that the ESG provides valid measures of positivity and negativity with a single response. In that bipolar valence measures can easily be derived from ESG ratings, the grid may prove valuable even when researchers are primarily interested in the bipolar valence dimension. In such studies, researchers can derive bipolar valence measures to test their focal hypotheses, then revisit the unipolar ratings in search of unexpected, but potentially illuminating, patterns of positivity and negativity.

Manuscript received 14 June 2006
Revised manuscript received 28 August 2007
Manuscript accepted 3 January 2008
First published online day/month/year

REFERENCES

Aitken, C. J., Siddle, D. A. T., & Lipp, O. V. (1999). The effects of threat and non-threat word lead stimuli on blink modification. *Psychophysiology*, *36*, 699–705.

Andrade, E. B., & Cohen, J. B. (2007). On the consumption of negative feelings. *Journal of Consumer Research*, *34*, 283–300.

Bradburn, N. M. (1969). *The structure of psychological well-being*. Oxford, UK: Aldine.

Bradley, M. M., & Lang, P. J. (1999a). *Affective Norms for English Words (ANEW): Instruction manual and affective ratings* (Technical report C-1). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Bradley, M. M., & Lang, P. J. (1999b). *International Affective Digitized Sounds (IADS): Stimuli, instruction manual and affective ratings* (Technical report B-2). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology*, *47*, 1191–1205.

Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H. S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology*, *50*, 260–268.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, *115*, 401–423.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review*, *1*, 3–25.

Center for the Study of Emotion and Attention (CSEA-NIMH). (1999). *The International Affective Picture System* [IAPS; photographic slides]. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Diener, E., & Emmons, R. A. (1984). The independence of positive and negative affect. *Journal of Personality and Social Psychology*, *47*, 1105–1117.

Duncan, S., Barrett, L. F., & Russell, J. A. (2006, January). *Bipolarity in continuous ratings of affective experience*. Poster presented at the annual meeting of the Society for Personality and Social Psychology, Palm Springs, CA.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, *53*, 712–717.

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2008). Mixed affective responses to music with conflicting cues. *Cognition and Emotion*, *22*, 327–352.

Ito, T. A., Cacioppo, J. T., & Lang, P. J. (1998). Eliciting affect using the International Affective Picture System: Bivariate evaluation and ambivalence. *Personality and Social Psychology Bulletin*, *24*, 855–879.

Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes*, *51*, 296–312.

Kaplan, K. J. (1972). On the attitude-indifference problem in attitude theory and measurement: A suggested modification of the semantic differential technique. *Psychological Bulletin*, *77*, 361–372.

Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review*, *97*, 377–395.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International Affective Picture System (IAPS): Instruction manual and affective ratings* (Tech. Rep. No. A-4). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*, 261–273.

Larsen, J. T., & McGraw, A. P. (2008). *"Midway between the two?": The case for mixed emotions.* Manuscript in preparation, Texas Tech University, Lubbock.

Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology*, *81*, 684–696.

Larsen, J. T., McGraw, A. P., Mellers, B. A., & Cacioppo, J. T. (2004). The agony of victory and thrill of defeat: Mixed emotional reactions to disappointing wins and relieving losses. *Psychological Science*, *15*, 325–330.

Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive affect and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, *40*, 776–785.

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science*, *8*, 423–429.

Norris, C. J., van Reekum, C. M., & Davidson, R. J. (2007). [*Evaluative reactions to emotional pictures as a function of age, race, and gender.*] Unpublished raw data, University of Wisconsin, Madison.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models.* Thousand Oaks, CA: Sage.

Reisenzein, R. (1995). On Oatley and Johnson-Laird's theory of emotion and hierarchical structures in the affective lexicon. *Cognition and Emotion*, *9*, 383–416.

Russell, J. A., & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin*, *125*, 3–30.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, *57*, 493–502.

Schimmack, U. (2001). Pleasure, displeasure, and mixed feelings: Are semantic opposites mutually exclusive? *Cognition and Emotion*, *15*, 81–97.

Schimmack, U. (2005). Response latencies of pleasure and displeasure ratings: Further evidence for mixed feelings. *Cognition and Emotion*, *19*, 671–691.

Scott, W. A. (1968). Attitude measurement. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 1 (2nd ed) (pp. 204–273)). London: Addison-Wesley.

Stark, R., Schienle, A., Walter, B., Kirsch, P., Blecker, C., Ott, U., et al. (2004). Hemodynamic effects of negative emotional pictures: A test-retest analysis. *Neuropsychobiology*, *50*, 108–119.
Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070.
Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*, 219–235.

APPENDIX

INSTRUCTIONS TO PARTICIPANTS AND GUIDELINES FOR USE

Instructions

Generic instructions for the computerised ESG are based on Russell et al.'s (1989) affect grid instructions; instructions may need to be modified for different types of stimuli. The experimenter shows participants the grid on a computer screen and explains:

> One way to describe your feelings about a stimulus is in terms of how positive and how negative you feel about it, as shown in the figure. It is in the form of a grid—a kind of map for feelings. The grid asks you two questions: Along the horizontal axis, it asks how positive you feel about the stimulus from "*not at all*" at the left to "*extremely*" at the right. Along the vertical axis, it asks how negative you feel from "*not at all*" on the bottom to "*extremely*" on top.
>
> If you feel positive, but not at all negative, move the mouse into one of these four cells at the bottom edge. [Throughout, the experimenter should move the mouse to highlight the appropriate cells.] The better you feel, the farther to the right you should go. On the other hand, if you feel negative but not at all positive, move the mouse into one of these four cells on the left edge. The worse you feel, the farther up you should go. If you feel neither positive nor negative, move the mouse into the cell in the bottom left. This indicates that you feel not at all positive and not at all negative. Finally, if you feel both positive and negative, move the mouse into one of the cells in the middle. The cell you select will depend on just how positive and just how negative you feel.
>
> The grid can be used to describe any pattern of positive and negative feelings. For example, you might feel quite a bit positive and not at all negative or somewhat negative and not at all positive. During the experiment, please place yourself in whichever cell best describes your feelings. After you've reached that cell, click to record your response.

Following the instructions, which take approximately 90 s to complete, participants were given approximately 30 s to get accustomed to moving around the grid with the mouse.

Guidelines for use

Pilot studies with the computerised grid indicated that ratings are not affected by whether (a) the origin (i.e., positive rating = 0; negative rating = 0) is located at the top left or bottom left or (b) positivity or negativity is located on the *x*-axis. Andrade and Cohen (in press) have replicated the latter effect. For simplicity, participants in Studies 1–4 completed the version shown in Figure 1.

Pilot studies with a paper-and-pencil version of the grid have revealed a somewhat different pattern of results. Specifically, when the origin is in the bottom left, ratings tend to be higher on the scale located on the *y*-axis. Whereas the cursor is placed at the origin prior to each trial in the computerised version of the grid, it is not possible to specify a default cell in the pencil-and-paper version. Participants may approach the pencil-and-paper grid as they approach written English: by starting at the top left corner. As a result, they may anchor their ratings on the cell at the top left corner. Thus, if the top left corner represents minimal positive affect and maximal negative affect (see Figure 1), negative ratings will be slightly higher than if the top left corner represents maximal positive affect and minimal negative affect. Consistent with this hypothesis, axis effects are not obtained when the origin is located in the top left. We therefore encourage researchers who use pencil-and-paper versions of the grid to place the origin in the top left.