Project Implicit technical report no. 2008-05

Comparison between the SPF and Evaluative Priming

Bar-Anan, Y. [1], Nosek, B. A.[1], & Vianello, M. [2] (2008)

[1]*Department of Psychology, University of Virginia*
[2]*Department of General Psychology, University of Padua*

baranan@virginia.edu

Available online 10 September 2008

Abstract

The Sorting Paired Features (SPF) task is an association measure that can be used to measure automatic attitudes. It has a few potential features that were formerly available mainly, if not solely, by Evaluative Priming, a well-established measure of automatic attitudes. Using the political domain, the present investigation compared the internal consistencies of the measures provided by these two tasks, the relationship between the two measures, and of each of them with self-reported attitudes. We found that the SPF is superior to Evaluative Priming in internal consistency, and is more strongly related to self-reported attitudes. In light of the SPF's psychometric qualities, we discuss the unique features of Evaluative Priming that make it particularly suitable for some research purposes.

This report is a supplement of the article that presents the Sorting Paired Features task (SPF) as a measure of association and automatic attitudes (Bar-Anan, Nosek, & Vianello, 2008; hereafter, "the main paper"). This manuscript reports a study that compared the SPF with a well-established and often-used measure of implicit attitudes, Evaluative Priming (EP, Fazio, et al., 1986, 1995). Our main purpose was to compare the psychometric features of the two measures, and also examine whether the two are related.

The SPF and EP are similar in a few features, detailed in the main paper. Of main interest in this report is that in both measures, the performance in each trial is affected by the processing of two stimuli: the prime and target in EP, and the two targets in the SPF. Each pair stimuli represent one of the four associations that affect performance in the task (e.g., *white-good*, *white-bad*, *black-good*, *black-bad* in a *black/white* attitude SPF). This may allow separating the trials by association and examining the effect of each association on performance separately.

The main paper found that a separate estimation of the association strength of each of the four associations may provide valuable information. For example, the results suggested that automatic pro-White biases, especially among White participants, are a function of both positive associations for Whites and negative associations for Blacks, the former more so than the latter. Another potentially interesting finding was what we called *good primacy:* in comparison to associations of categories with negative valence, associations with positive valence showed higher internal consistency and stronger relationship with other attitude measures. Although it is uncommon, the same four separate measures can be computed from EP data as well. This study examined this possibility and compared the four measures across tasks in the domain of political attitudes.

Method

Participants and materials. Four hundred and one American volunteers (64% females; Mean age = 30.15; SD = 13.76) at Project Implicit (https://implicit.harvard.edu; see Nosek, 2005) completed at least one measure.

Stimuli for both EP and SPF were images of 10 American Republican and Democrat political figures (5 of each party).

Measures. The *SPF* procedure was identical to the one used in Study 2 of the main paper (3 blocks of 40 trials each).

EP. In the first block (28 trials), participants categorized the valence of 14 positive and 14 negative adjectives. This was also the main task in blocks 2-4 (40 trials each), but a secondary task was to pay attention to images that appeared before each adjective, in preparation for a recognition test. The images were the same politicians images used in the SPF. The Images appeared for 200ms, followed by an inter stimulus interval of 50ms until the display of the target word. A response deadline was used: If participants did not respond in 1500ms or less, the trial ended, and participants saw a message urging them to answer more quickly. The next priming trial was initiated after 500ms. Block 5 was a recognition test in which participants categorized some images as *old* or *new*.

Explicit Attitudes. Participants were also asked to report their feelings toward each candidate and each political party on a thermometer scale (0 to 8). They also reported their voting intention in the upcoming 2008 election (the study was conducted at the end of August 2008).

Procedure. The order of the measures was counterbalanced across participants.

Results

Scores. Both SPF and EP measures were scored using the D scoring algorithm (Greenwald, Nosek, & Banaji, 2003). Error trials were included in the SPF analysis because participants could correct their response (the trial latency was the latency of the correct response). Error trials were not included in the EP analysis because EP does not allow correction. The error-rate for EP was 16%. Therefore, the average number of informative trials in the SPF was 112, in comparison to only 98 in the EP. This might be a procedural advantage of the SPF, and it cannot be used in the EP unless the response deadline is removed. Information on the SPF scores when error trials are removed (mean error rate = 14%) appear at Table 4.

Table 1 presents the means, standard deviations and internal consistency of the SPF and EP measures. It appears that both measures had a lower internal consistency when the measure was administrated first. This reduction was close to significance only for the EP *Democrats+bad*, EP *Republicans+*bad and overall EP preference, Fischer *z*s = 1.87, 1.96 ,1.93, *p*s = .060, .049 .052, respectively. This probably reflected the importance of category accessibility. Each measure may have increased the accessibility of the categories, and therefore the categories were more influential on performance in the subsequent measure. As expected, this effect was much stronger for EP because, unlike the SPF, the categories are not mentioned in the task, so previous mentioning of them is likely to have a more substantial effect on accessibility.

Table 1. Means, standard deviations and internal consistency

	Mean	*SD*	*IC (all participants)*	*IC (when the measure was first)*
SPF Dems+good	.02	.22	.60	.49
SPF Dems+bad	-.01	.17	.44	.32
SPF Reps+good	-.02	.21	.54	.48
SPF Reps+bad	.01	.18	.39	.37
SPF Preference	.05	.51	.65	.59
EP Dems+good	.06**	.19	.25	.03
EP Dems+bad	-.08**	.19	.23	.17
EP Reps+good	.03**	.21	.54	.44
EP Reps+bad	-.03**	.18	.25	.02
EP Preference	.08**	.48	.33	.11

Notes. IC = Internal Consistency. n for the all participants condition: SPF = 353, EP = 353; n for the measured-first condition: SPF = 113, EP = 111.

Most of the SPF scores showed superior internal consistency over their parallel EP scores. The only association that did not show this difference was *Republicans-good*, as a result of the relatively high consistency found in EP for this association. Additionally, in the "automatic-measure-first" condition (last column in Table 1), there was no significant difference between the internal consistencies of the *Democrats-bad* EP and SPF association scores. All the others were significantly different, *ps* < .05.

Explicit attitudes and voting intentions. Because the correlation between the preference item (M = 4.9, SD = 1.97) and the difference between the two thermometer items (M = 2.24, SD = 4.46) was $r(434) = .91$, $p < .0001$, the explicit attitude index was the mean of their standardized values. 71% reported intention to vote for the Democratic candidate in the 2008 election.

Relationship between the measures. As illustrated in Table 2, the separate scores of the SPF and EP did not show strong evidence of divergent and convergent validity. That is, the diagonals in Table 2 did not display the stronger correlations. This may be explained by the internal consistencies (see Table 1): Because the scores for associations with bad were less consistent, there were lower chances for strong correlations between two scores of bad associations than between a bad association score and a good association score.

A comparison of the average correlation magnitude of the same-association correlations along the diagonals and the different-association correlations along the off-diagonals found a slight advantage for the

on-diagonal correlations over the off-diagonals among all participants and the SPF-first participants, but not among the EP-first participants. This might be considered as a slight evidence for convergent validity. Future investigations with various attitude domains and method that control for reliability differences (e.g., structural equation modeling) may help to test this issue further.

Table 2. SPF/EP correlation, separate measures

All Participants	***EP Dems+good***	***EP Dems+bad***	***EP Reps+good***	***EP Reps+bad***
SPF Dems+good	.21**	-.18**	-.21**	.19**
SPF Dems+bad	-.04	.12*	.06	-.14**
SPF Reps+good	-.21**	.07	.30**	-.18**
SPF Reps+bad	.01	.03	-.14**	.10
EP First	EP Dems+good	EP Dems+bad	EP Reps+good	EP Reps+bad
SPF Dems+good	.25**	-.26**	-.13	.17
SPF Dems+bad	.08	.04	.00	-.13
SPF Reps+good	-.24*	.15	.23*	-.15
SPF Reps+bad	-.11	.14	-.07	.03
SPF First	EP Dems+good	EP Dems+bad	EP Reps+good	EP Reps+bad
SPF Dems+good	.24**	-.24**	-.20*	.17
SPF Dems+bad	-.24**	.17	.06	.03
SPF Reps+good	-.18*	.10	.37**	-.26**
SPF Reps+bad	.13	.03	-.25**	.08

*Notes. * p < .05; ** p < .01.*

We examined the relationship between the automatic measures and self-report by focusing on the conditions in which the relevant automatic measure was completed first and the self-report measure was completed second. These seems to be the appropriate conditions for comparison because (a) EP must be completed first because the other measures increase the accessibility of the categories, and (b) the non-relevant automatic measure must be completed after the self-report, to make sure that it does not affect the self-report.

As detailed in Table 3, the SPF scores were superior to the EP scores. Out of 20 examined correlations between SPF scores and explicit indexes, 15 were above .24 and significant. In comparison, only 2 correlations were above .24 and significant among the EP scores. The relationship between SPF preference and the explicit preference was significantly stronger than the relationship of the EP preference and the explicit preference, Fischer $z = 1.96$, $p = .05$.

The correlations across all participants, ignoring the task-order restrictions, were rs(353) = .49, .38, .34 for SPF/Explicit, EP/Explicit and SPF/EP, respectively, ps < .001. The difference between the SPF/Explicit and the EP/Explicit correlations was significant, Williams $t = 2.1$, $p = .03$ (Steiger, 1980).

Table 3. Correlation between the SPF and EP scores and explicit measures.

	Explicit preference	*Explicit Dems*	*Explicit Reps*	*Preference by items*	*Voting intentions*
SPF Dems+good	.47**	.41**	-.45**	.51**	.36**
SPF Dems+bad	-.28*	-.24*	.26*	-.25*	-.12
SPF Reps+good	-.37**	-.32**	.36**	.45**	-.34**
SPF Reps+bad	.14	.12	-.14	.16	.09
SPF Preference	.51**	.44**	-.49**	.56**	.37**
EP Dems+good	.22	.09	.-29*	.30*	.13
EP Dems+bad	-.17*	-.16	.15	-.25*	-.27*
EP Reps+good	-.04	-.05	.01	-.07	.06
EP Reps+bad	.04	.15	.06	.10	.11
EP Preference	.20	.18	-.16	.30*	.18

*Notes. * p < .05; ** p < .01. Across people who first completed the relevant automatic measure and then the self-report (SPF-first for the SPF scores, and EP-first for EP scores; ns = 72, 59, respectively). Preference by items was computed from the rating of the individuals politicians used in the tasks.*

Finally, as detailed in Table 4, when the SPF scores were computed without error trials, their psychometric indices were naturally reduced in quality, but were still numerically better than the parallel indices among EP scores.

Table 4. The psychometric indices of the SPF scores without error trials.

	Explicit preference (when SPf first and explicit second)	*Intend to Vote Dems (when SPF first and explicit second)*	*IC (all participants)*	*IC (when SPF was first)*
n	77	77	355	114
SPF Dems+good	.45**	.32**	.53	.42
SPF Dems+bad	-.21	-.11	.40	.35
SPF Reps+good	-.32**	-.26*	.43	.42
SPF Reps+bad	.09	.13	.30	.09
SPF Preference	.42**	.33**	.56	.46

*Notes. * p < .05; ** p < .01.*

Good Primacy. In Tables 1-4, the internal consistencies and correlations of the *Democrats+good* and *Republicans+good* association scores were always larger than those of *Democrats+bad* and *Republicans+bad*. This is a replication of the good-primacy effect reported in the main paper. This was not always the case with the EP scores, but perhaps their lower internal consistency requires using a larger sample in order to uncover good-primacy. Table 5 presents the relationship between the four measures of each of EP and the SPF with the preference measures,

across all participants. The good-primacy is apparent for the SPF scores but not for EP.

Table 5. Correlation between the SPF and EP scores and explicit measures.

	Explicit preference	*Preference by items*	*Voting intentions*	*other automatic measure preference*
SPF Dems+good	.43**	.48**	.35**	.33**
SPF Dems+bad	-.21**	-.23**	-.13*	-.15**
SPF Reps+good	-.43**	-.47**	-.35**	-.31**
SPF Reps+bad	-.18**	-.17**	-.11*	.09
EP Dems+good	.25**	.27**	.21**	.19**
EP Dems+bad	-.19**	-.20**	-.22**	-.13*
EP Reps+good	-.26**	-.29**	-.17**	-.28**
EP Reps+bad	.23**	.25**	.21**	.24**

*Notes. * p < .05; ** p < .01. n= 353.*

The *good-primacy* effect was found in each of the SPF studies presented in the main paper and the supplements. A similar effect is found with numerous studies that we conducted with the Brief-IAT (Sriram & Greenwald, 2008). Unkelbach and colleagues (2008) have recently reported that positive and negative attributes have different effect on performance in EP. They found that positive words are processed faster than negative words and that the positive-positive facilitation in EP is larger than the negative-negative facilitation. This does not necessarily affect the psychometric indices that are the focus of the good-primacy effect. Therefore, at the time being, we cannot say that the good-primacy effect is related to the EP findings, and we cannot say whether the density hypothesis suggested by Unkelbach and colleagues (2008) is the reason for the good-primacy effect.

Table 6. Intercorrelations within each measure

	Dems+bad	*Reps+good*	*Reps+bad*
SPF *Dems+good*	-.40**	-.60**	-.17**
SPF *Dems+bad*	--	-.11*	-.33**
SPF *Reps+good*	--	--	-.32**
EP *Dems+good*	-.51**	-.27**	-.22**
EP *Dems+bad*	--	-.27**	-.17**
EP *Reps+good*	--	--	-.52**

*Notes. * p < .05; ** p < .01. n= 353.*

Intercorrelations. Table 6 presents the intercorrelations among the EP and SPF scores. Note that the negative correlations are expected because the scores are algebraically dependent – a comparison between the correlations' magnitude is the primary meaningful data. The patterns of intercorrelations for each measure seem different. In the SPF, the strongest relationship is between the two associations with good, whereas in EP, the strongest relationships are between the two associations of each category. The EP intercorrelations among participants who completed this measure first reveals a similar pattern with the exception that the *Republicans-Bad* and *Democrats-Bad* associations show correlation as strong as the correlations between the same-category associations. More data is needed before speculation or assumptions could be drawn from that.

Discussion

The SPF's psychometric qualities presented in the main paper were superior to those usually found with EP. The present study supported this conclusion by comparing the two measures in the same study. At the same time, in most of our studies, the SPF usually showed psychometric qualities inferior to the IAT and the AMP (in studies that included the IAT, this inferiority was not significant, but the accumulative numerical inferiority is convincing). Therefore, researchers are recommended to consider a host of factors when they select a measure for their research.

As suggested in the main paper, the SPF seems appropriate for researchers who are interested in the potential additional information that the IAT and its variants failed to provide in past research. This information is mainly the separate association measures and measures of non-focal associations. Also, researchers who have strong reasons to avoid the order-effects found with the IAT and its variants might consider using the SPF. Finally, the SPF may prove a valuable tool in examining associations. For instance, the investigation of the good-primacy effect may prove useful in understanding mental associations in general, and may provide information to improve our understanding and usage of other measures.

The results of the present study do not suggest that researchers should stop using EP. First, it should be noted that we used an equal number of trials for the SPF and EP to allow comparison. However, experiments in our lab and personal communications with other researchers suggest that EP psychometric qualities improve at the end of the task compared to the beginning. This may be a function of the prime categories becoming more accessible as trials accumulate and participants recognize a pattern in the nature of the primes. .

Second, EP does not require any explicit definition of the categories. The SPF explicitly defines focal categories (although it also allows the measurement of associations with non-focal categories). This could be a negative feature for some research applications. Note that no explicit definition of the categories is also a feature of the AMP (Payne et. al, 2005) and the EAST (De-Houwer, 2003). Third, EP can measure associations with 1 to many categories in a single

response block. The current procedure of the SPF measures four and only four associations.

In conclusion, our study found that the SPF shows better internal consistencies and relationship with other measures in comparison to EP. We also found that good-primacy—the psychometric superiority of associations with good over associations with bad—is very replicable in the SPF, but was not found in EP. We suggest that researchers should consider using the SPF as a measure of associations, with or instead-of other measures, according to their specific research needs. Beyond its potential usage as a measure, the SPF is a cognitive task that may advance our study of the psychological processes that affect categorization and processing of symbolic stimuli.

References

Bar-Anan, Y., Nosek, B. A., & Vianello, M. (2007). The Sorting Paired Features task: A measure of association strengths. Unpublished manuscript.

De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology, 50*, 77-85.

Fazio RH , Sanbonmatsu DM , Powell MC , Kardes FR . (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229 – 38

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C.J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013–1027.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464–1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85(2), 197-216.

Payne, B.K., Cheng, C. M., Govorun, O., & Stewart, B. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, 89, 277-293.

Unkelbach, C., Fiedler, K., Bayer, M., Stegmuller, M., & Danner, D. (2008). Why positive information is processed faster: The density hypothesis. *Journal of Personality and Social Psychology, 95*, 36-49.

Sriram, N. & Greenwald, A. G. (2008). The Brief Implicit Association Test. Unpublished manuscript.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245-251.

Teige-Mocigemba S., Klauer, K. C. (2008). 'Automatic' evaluation? Strategic effects on affective priming. *Journal of Experimental Social Psychology, 44*, 1414-1417.