Project Implicit technical report no. 2007-01

Reliability of the SPF: Test-retest and internal consistency

Michelangelo Vianello[1], Yoav Bar-Anan[2], & Brian A. Nosek[2] (2007)

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

michelangelo.vianello@unipd.it

Available online 29 November 2007

Abstract

This study explores test-retest reliability and internal consistency of the SPF. Since the stability of implicit attitudes and preferences is still under scrutiny, we adopted the IAT as a comparison standard. We chose a three-wave longitudinal study design in which sixty participants were asked to return after a week (Time 2) and after a month (Time 3) from the first measurement occasion (Time 1). The SPF showed relatively high internal consistencies (mean α= .71) across measurement occasions and good test-retest reliabilities. The mean test-retest correlation observed with the SPF (mean *r*=.42) was indeed not significantly different from that obtained with the IAT (mean *r*=.64).

This study is a supplement to Bar-Anan, Nosek, & Vianello (2007) and was conceived to explore the test-retest reliability of the SPF. A main problem in test-retest validations is the assumption of stability of the latent trait, which directly affect both predictions and interpretations of test-retest correlations. This problem is quite evident, for instance, in test-retest validation of state and trait anxiety scales. While we suppose that a time-1 series of scores obtained by our trait anxiety scale will correlate highly with scores obtained by the same participants at the same scale in time 2 (let's say after a month), we expect that our state anxiety score will change according to the specific situation in which the person is (e.g. it should be quite high just before an examination or a talk and rather low just after these events). This problem is exacerbated in the study of automatic associations, since the stability of implicit attitudes, preferences and stereotypes is still under scrutiny. An ideal way of dealing with this problem would be to know whether an implicit attitude or preference is stable over time. In the absence of such knowledge, we built our theoretical predictions on the basis of previous test-retest correlations of a reliable measure of implicit attitudes, the IAT (Greenwald et al. 1998). We included the IAT in this study as a basis of comparison with the test-retest correlations for the SPF.

Method

Participants and materials. Sixty students (41 females) participated in this experiment on a voluntary basis. Participants were asked to return after a week (Time 2) and after a month (Time 3) from the first measurement occasion (Time 1).

Stimuli were both words and images. We used words to represent elements of good and bad categories and 32 images to represent two attitude objects: sweet and salty foods (16 stimuli each). The list of good and bad words and four examples of sweet and salty images are reported in the appendix. For both implicit measures the intertrial interval was set to 150 ms. and participants had to correct their response in case of errors.

Measures. The *SPF* was similar to the one used in Study 2 (of Bar-Anan, Nosek, & Vianello, 2007), with the following differences: 1) there were 24 practice trials, and they provided stimuli of the same four categories used in critical blocks (i.e. good words, bad words, salty foods and sweet foods, 2) the three critical blocks were 21 trials each, and 3) participants categorized items into the four corners using a touch screen rather than a keyboard. The *IAT* followed the original IAT design (Greenwald et. al, 1998), but with only one 56-trial block for each combined pairing

condition. The order of the test blocks was counterbalanced.

Self-report: participants rated each stimulus on a 7-point Likert scale (1=very bad; 6=very good). Responses were then summed to form overall explicit scores toward sweet and salty foods.

Results

Both SPF and IAT measures were scored using the D scoring algorithm (Greenwald, Nosek, & Banaji, 2003). For all preference scores, positive numbers indicate preference for sweet over salty food.

Table 1. *Means and standard deviations of implicit and explicit measures across measurement occasions.*

	Time 1		*Time 2*		*Time 3*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
SPF sweets+bad	.05	.30	-.02	.32	-.07	.31
SPF salty+bad	.01[a]	.41	.16[a]	.27	.14	.30
SPF sweets+good	.18	.27	.09	.31	.15	.25
SPF salty+good	-.24	.26	-.23	.29	-.22	.29
SPF sweets	-.04[b]	.64	.18[b]	.52	.20	.54
SPF salty	-.41	.40	-.31	.52	-.37	.45
SPF	.37	.76	.49	.70	.57	.74
IAT	.73	.41	.73	.36	.65	.42
Explicit sweets	66.1	11.8	65.5	11.9	64.2	11.6
Explicit salty	63.8	11.4	65.2	11.5	63.8	12.3

Notes:
a) SPF sweets is a D score calculated comparing sweets+good and sweets+bad associations. SPF salty is a D score calculated comparing salty+good and salty+bad associations. SPF is a D score calculated as the combination of all four associations indicating a relative preference for sweet over salty foods.
b) "a" and "b" identify pairs of different mean values (p<.05)

Relations among measures. Across occasions the SPF sweet-salty preference score was weakly related with sweet-salty IAT preferences (r = .14) and self-reported sweet-salty preferences (r = .06). Likewise, the self-report and IAT measures were relatively weakly related on average (r = .22).

Internal consistency of implicit measures. We estimated internal consistencies computing an SPF score for each of the three blocks provided and then computing a Cronbach's alpha on the three measures. Reliability for the IAT was computed correlating the scores of the first 20 trials with the last 36 trials. Internal consistencies of the SPF were relatively high compared to most implicit measures (α= .71) and of comparable magnitude to those often observed with the IAT in this case (α= .69) and in general (Nosek, Greenwald, & Banaji, 2006), or the AMP (Payne, Cheng, Govorun & Stewart, 2006).

Table 2. *SPF and IAT Internal Consistencies.*

	Time 1	*Time 2*	*Time 3*	*Mean across times*
SPF sweets+bad	.84	.69	.75	.76
SPF salty+bad	.73	.78	.83	.78
SPF sweets+good	.67	82	.89	.79
SPF salty+good	.40	.85	.82	.69
SPF sweets	.52	.78	.83	.71
SPF salty	.24	.78	.33	.45
SPF	.75	.71	.86	.77
IAT	.67	.72	.68	.69
Explicit –sweet foods	.89	.92	.91	.91
Explicit –salty foods	.86	.88	.90	.88

Test-Retest reliability. SPF association measures did not show mean differences over time (Table 1), with the exception of the sweets-good association. This association was stronger at Time 2 compared to Time 1 ($F_{(1,57)}$=7.89, p=.007, η^2=.122).

Table 3. *SPF and IAT Test-Retest Correlations.*

Sweets vs. Salty foods	*r with time 2* (7 days)	*r with time 3* (23 and 30 days)	disattenuated *r*s	
SPF sweets+bad Time 1	.42**	.51**	.56	.71
SPF sweets+bad Time 2		.64**		.89
SPF salty+bad Time 1	.42**	.37**	.56	.46
SPF salty+bad Time 2		.56**		.70
SPF sweets+good Time 1	.29*	.10	.39	.11
SPF sweets+good Time 2		.35**		.41
SPF salty+good Time 1	.53**	.39**	.91	.47
SPF salty+good Time 2		.51**		.61
Mean	*.42*		*.56*	
SPF sweets Time 1	.18	.01	.28	.02
SPF sweets Time 2		.36**		.44
SPF salty Time 1	.56**	.05	1.28	.11
SPF salty Time 2		.12		.23
Mean	*.23*		*.39*	
SPF Time 1	.16	-.16	.21	-.21
SPF Time 2		.20		.26
Mean	*.07*		*.09*	
IAT Time 1	.64**	.65**	.92	.93
IAT Time 2		.64**		.91
Mean	*.64*		*.92*	
Explicit sweets Time 1	.77**	.76**	.85	.84
Explicit sweets Time 2		.83**		.91
Explicit salty Time 1	.83**	.72**	.95	.82
Explicit salty Time 2		.83**		.93
Mean	*.76*		*.89*	

Note. * p < .05, ** p < .01.

No difference was observed between Time 2 and Time 3 ($F_{(1,57)}$ =.45, p=.50, η^2=.008). Mean IAT effect sizes did not change over time ($F_{(1,57)}$ =1.73, p=.18, η^2=.029). The mean SPF test-retest correlation was .42 (Table 3) for the individual associations. This correlation was not significantly different from that obtained with the IAT (r = .64). And was lower for the combined sweets, salty, or relative measure calculations (mean r's = .23, .07).

Discussion

This study provides evidences that the SPF has internal consistency that is comparable to the IAT and test-retest reliability that is somewhat lower. The mean internal consistency of the SPF (α=.71) was higher than some implicit measures and comparable to other observations with the IAT and the AMP. A potentially important procedural difference between the present study and previous research is that participants responded by means of a touch-screen instead of a keyboard. It is unknown whether this has an impact on reliability. Taken together, these results suggest the SPF is a reliable measure of implicit preferences.

References

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

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.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2006). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes* (pp. 265-292). Psychology Press.

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.

Appendix

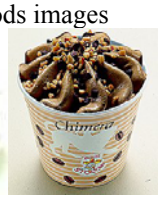
List of good words

Good
Pleasure
Tasty
Pleasant
Nice
Heaven
Wonderful
Marvellous
Rich
Better
Favour
Sublime
Lofty
Excellent
Funny
Attractive

List of bad words

Bad
Nasty
Stink
Disgusting
Reek
Awful
Monstrous
Rotten
Unpleasant
Nausea
Disgust
Terrible
Hell
Worst
Unpleasant
Poor

Examples of sweet foods images



Examples of salty foods images