The Harris-Westin's Index of General Concern About Privacy: An Attempted Conceptual Replication.

A summary of a presentation prepared for the GPD International Survey Research Workshop, Kingston, Ontario, November 17-18, 2006.

Stephen T. Margulis, Jennifer A. Pope, and Aaron Lowen

Seidman College of Business, Grand Valley State University

Grand Rapids, Michigan, U.S.A.

Notes

Please address all queries about this summary to Dr. Stephen T. Margulis at margulis@gvsu.edu.

The authors wish to thank Emily Smith and William Rogers for their help.

The Harris-Westin Index of General Concern About Privacy

Alan Westin, among the most influential theorists about privacy, uses a classificatory approach as a central feature of his theorizing (see Margulis, 2003, for a description and review). A prime example is his classification of privacy's states and functions. Another example is a classificatory approach to an individual difference measure of privacy concern that he and Louis Harris and Associates introduced in a 1990 Harris-Equifax poll of privacy opinions of U.S. residents (Harris & Westin, 1991). Harris and Westin called their measure an Index of General Concern About Privacy.

The index is based on four items. For each item, one specific response is used to indicate high concern. These four responses are used to generate a three-fold classification of privacy concerns. One question asks respondents about their concern "about threats to their personal privacy today"; the high concern response is "very concerned". A second question asks respondents whether they agree "that business organizations seek excessively personal information from consumers"; the high concern response is "strongly agree". A third item asks respondents whether they agree "that the [U.S.] Federal government since Watergate is still invading the citizen's privacy"; again, the high concern response is "strongly agree". The fourth item asks respondents whether they agree that "consumers have lost all control over how personal information about them is circulated and used by companies"; the high concern response is "agree". Whether "agree" means "strongly agree," as in the aforementioned items, or both "strongly agree" and "Agree somewhat strongly" is unclear (Harris & Westin, 1991, pp. 6, 30). Presumably, access to the 1990 report, on which the 1991 discussion is based, would resolve this issue.

Harris and Westin classify respondents into three categories based on how often respondents express high concern. Those classified as High Concern respondents have expressed high concern on three or four of these items; Moderate Concern respondents have expressed high concern on two of the four items; Low Concern respondents have expressed high concern on one or none of the four items (Harris & Westin, 1991).

When Harris and Westin examined the relationship between their three privacy concern groups and the groups' responses to other polling item on business and governmental aspects of personal privacy, Harris and Westin found that "[i]n a majority of these [analyses], the index worked as a direct predictor – people in the High General Concern About Privacy group were the most concerned about privacy interests on specific questions; people in the Moderate Concern group were next in their privacy concern; and people in the Low Concern group were the lowest in privacy concern" (Harris & Westin, 1991, p. 6).

Westin (see Harris & Westin, 1990, p. 6) appears to equate Privacy Fundamentalists with the High Concern group. Privacy Fundamentalist "rejected consumer-benefit or societal-protection claims for data uses and sought legal-regulatory privacy measures" (Westin, 2003, p. 445). Westin appears to equate the Privacy Unconcerned with the Low Concern group. The Privacy Unconcerned "were generally ready to supply their personal information to business and government and rejected what was seen as too much privacy fuss" (Westin, 2003, p, 445). Westin appears to equate Privacy Pragmatists with the Moderate Concern group. Westin regards Privacy Pragmatists as holding a more balanced view of privacy because they "examined the benefits to them or society of the data collection and use, wanted to know the privacy

risks and how organizations proposed to control those, and then decided whether to trust the organization or seek legal oversight" (Westin, 2003, p. 445). The pragmatic position is consistent with Westin's own position on privacy (cf. Westin, 1970).

The Harris-Westin Index has been used in subsequent Harris polls (Smith, 2006). Based on trend data, Westin (2003) concluded that the percent of the U.S. population who express low concern has been dropping in recent years, privacy fundamentalists have been increasing, and privacy pragmatists have remained the majority (p. 446). Table 1 compares the percentage of respondents in each classification in 1990, when the index was created, and 2003, the latest source of data available to the authors (Smith, 2006).

Table 1. A Comparison of the Percent of Privacy Fundamentalist, Privacy Pragmatists, and Privacy Unconcerned in 1990 and 2003.

	1990	2003
Privacy Fundamentalists	25%	26%
Privacy Pragmatists	57%	64%
Privacy Unconcerned	18%	10%

A Comparison of Harris-Westin and GPD Privacy Concern Index Items

Our proposal to the GPD Project was to examine Westin's ideas about privacy concern using multi-nation data from the GPD Project. We hoped to create an index of privacy concern and, if that was possible, to examine its correlates in GPD survey data. By way of summary, we found a general concern factor in our factor analysis. But nothing else turned out as expected and we don't know why.

Because the GPD Project survey questionnaire did not include the Harris-Westin privacy concern items, we could not directly replicate the Harris-Westin measure. However, we believed that the GPD questionnaire included items that could provide a

basis for a conceptual replication of the index. That is, there are GPD questionnaire items that reasonably fit the items/issues that are the basis for the Harris-Westin index. Smith (2006) reached a similar conclusion.

The GPD questionnaire items we believe are comparable to each of the four items in the Harris-Westin Index of General Concern about Privacy follow. We also include our interpretation of the degree of general concern implied by the GPD items' response alternatives.

Harris-Westin: "How concerned are you about threats to your personal privacy in America today?"

GPD item 6. "What level of trust do you have that private companies, such as banks, credit card companies and places where you shop, will protect your personal information?"

For us, this item's response alternatives "Fairly low level of trust" and "Very low level of trust" represent high concern. "Reasonably high level of trust" represents intermediate concern. "Very high level of trust" represents low concern.

GPD item 10 addresses predicted governmental efforts to protect the privacy of personal information on national identification cards, if these cards were issued. The item reads: "In order to put national ID [sic] cards into use, the government would need to have a national database containing personal information on all citizens. This information could include address, gender, race, and tax information. How effective do you feel efforts to protect this type of information from disclosure would be?"

For us, the response alternative "Not at all effective" and "Not very effective" represent high concern. "Somewhat effective" represents an intermediate level of concern. "Very effective" represents a low level of concern.

GPD item 11. "When it comes to privacy, how worried are you about providing personal information on websites, such as your name, address, date of birth, and gender?" For us, the response alternative "Very worried" represents high concern. "Somewhat worried" and "Not very worried" represent intermediate levels of concern. "Not worried at all" represents a low level of concern.

We would have included **GPD item 4,** "To what extent do you believe laws are effective at protecting your personal information that is held by government departments and private companies?" However, a filter item, GPD item 3, which assessed knowledge privacy laws, substantially reduced (by 46%) the number of possible respondents.

Harris-Westin. "Consumers have lost all control over how personal information about them is circulated and used by companies."

GPD item 2. "To what extent do you have a say in what happens to your personal information?"

Response alternative "No say" represents high concern. "Some say" represents a somewhat lower level of concern. "Complete say" and "A lot of say" are even lower levels of concern.

Harris-Westin. The item asks respondents whether they agree that business organizations seek excessively personal information from consumers.

Harris and Westin (1991) do not provide the actual wording of this item or the following item.

We agreed that no **GPD** items directly addressed this topic.

Harris-Westin. The items ask respondents whether they agree the U.S. government since Watergate is still invading the citizen's privacy.

GPD item 17 is about whether laws aimed at protecting national security are intrusive upon personal privacy. "The government of [*insert respondent's nation*] has enacted laws aimed at protecting national security. To what extent do you believe laws aimed at protecting national security are intrusive upon personal privacy?"

The response alternative "Highly intrusive" represents high concern. "Somewhat intrusive" represents an intermediate level of concern. "Not very intrusive" and "Not intrusive at all" represent lower levels of concern.

GPD item 23. "To what extent is your privacy respected by airport and customs officials when traveling by airplane?"

The response alternative "Not respected at all" represents high concern. "Somewhat respected" represents an intermediate level of concern. "A lot of respect" and "Completely respected" represent low levels of concern.

Given presumably comparable items, we next addressed whether to follow the Harris-Westin classificatory approach to privacy concern. We did not follow the Harris-Westin approach for several reasons. First, their approach reduces ordinal or perhaps interval response scale data to nominal data, thereby eliminating potentially useful

behavioral variance in the responses to these items. Second, we found that our comparable items might be multi-factorial. (We describe the factor analyses below.) If that was the case, then combining items across factors into a single scale measuring a single concept would have been methodologically problematic [W. Rogers, personal communication, September 22, 2006]. Third, we did not how Harris and Westin treated respondents who had answered any of their four items with a "Not Sure" response. Harris and Westin (1991) report Not Sure rates for two of the four items in their index. For the item on concern about threats to your privacy, the Not Sure rate was one percent in 1990 and 1991 (Harris and Westin, 1991, table 3, p. 6). For the item on loss of control over personal information, the Not Sure rates were three and four percent, respectively, in the 1990 and 1991 polls (Harris and Westin, 1991, table 6, p. 9). We believe that respondents with a Not Sure response to any of the four general concern items would have been eliminated. Otherwise, the classification rules have the potential of misclassifying respondents. Fourth, we did not know Harris and Westin's explanation for choosing the cut-offs for their three levels of concern (e.g., why are three high concern responses considered High Concern but two High-Concern responses are considered Moderate Concern?). For these reasons, we opted for a parametric approach to scaling privacy concern.

The Use of U.S. and Canadian Data

Before discussing, in detail, the GPD privacy concern index, we wish to explain why we took advantage of the availability of U. S. and Canadian data. We chose U.S. data because the Harris-Westin polls are based on U.S. samples. We chose Canadian data because, sociopolitical differences notwithstanding, Americans and Canadians are

culturally similar but not that similar. Pope (2003), using the cultural distance index of Barkema and Vermuelen (1997), reports that Canada and the U.S. have a 0.272 cultural distance ranking. By comparison, the United Kingdom (with a cultural distance rank of 0.110) and Australia (with a cultural distance rank of 0.021) are more culturally similar to the United States than Canada is. Given the degree of Canadian-U.S. cultural difference, we examined U.S. and Canadian data separately. The sizes of the initial samples were 1,001 respondents in the Canadian sample and 1,000 respondents in the U.S. sample.

One GPD Privacy Factor or Two?

Before constructing a measure of privacy concern, we decided to factor analyze our six GPD privacy concern items. The anti-image correlation matrix did not have any correlations above 0.3 showing that this data is appropriate for factor analysis. The factor analyses excluded all respondents, in both the U.S. and Canadian samples, who answered with a Don't Know/Not Sure to any of the six items. We also recoded the six items so "4" represented high concern on all of the items. We used a principal components analysis extraction method to generate an un-rotated component matrix and a Varimax Rotation, using a Direct Oblimin Rotation with Kaiser Normalization, to generate a rotated solution. We report the results of the principal components analysis because it was more straightforward to interpret than the rotated solution. Table 2, on the next page, summarizes the component matrix. All factors loading above 0.5 are in bold. The table includes the results of two tests, the Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett's test of sphericity. They are different ways to ensure that factor analysis is the appropriate methodology, similar to the anti-image correlation matrix (Factor Analysis, 2005).

Table 2. Factor Loadings for U.S. and Canadian Samples

Question	Canada-Factor 1	Canada-Factor 2	USA-Factor 1	USA-Factor 2
Question 2	0.454	**-0.707**	0.431	**-0.633**
Question 6	**0.719**	-0.182	**0.612**	-0.032
Question 10	**0.695**	0.030	**0.715**	-0.029
Question 11	0.355	**0.559**	0.395	**0.752**
Question 17	**0.514**	0.471	**0.515**	-0.028
Question 23	**0.551**	-0.018	**0.602**	-0.015
KMO [a]	0.669		0.702	
Bartlett's [b]	χ^2=228.9	p=0.000	χ^2=207.79	p=0.000

a. Kaiser-Meyer-Olkin Measure of Sampling Adequacy.
b. Bartlett's Test of Sphericity

We found one, perhaps two, factors in both the U.S. and Canadian samples. The factor loadings in the Canadian and U.S. data are quite similar. We say "perhaps two" because, although the eigenvalues of both factors are greater than 1.0 for the Canadian sample, the eigenvalue of Factor 2 in the U.S. sample is less than 1.0 (0.969). Our experts strongly disagree about the justifiability of including a factor with an eigenvalue less than 1.0 and whose two keys items (items in bold in Table 2) do not load sufficiently differently on the two factors.

The first factor/component has four items (GPD items 6, 10, 17, and 23), each loading at .5 or higher on the factor. Several of these items focus on general (personal) concerns about the influence of external institutions or forces on the privacy of personal information. Because of that emphasis on general (personal) concerns, this factor is, potentially, a conceptual replicate of the Harris-Westin index. Rather than name this factor, we call it Factor 1.

The problematic second factor, called Factor 2, has two items; one loads positively (GPD item 11) and one loads negatively (GPD item 2) on the factor. Both items also load noticeably on Factor 1. These items appear to be less a measure of

concern than a counter to general concern. GPD item 2 most clearly is internally focused on whether the person is able to protect the privacy of his or her personal information. However, because we do not agree on the interpretation of GPD item 6 (in Factor 1), on trust in external institutions to protect personal information, we cannot agree on whether item 6 has the same focus as item 11 (in Factor 2). As a result, we have unresolved issues about the appropriate interpretation of Factor 2.

For purposes of data analysis, we created two parametric measures, one for each factor, using factor scores based on our factor analytic results. We used the Anderson-Rubin method to create factor scores because it ensures the orthogonality of the estimated factors (Factor Analysis, 2005).

Correlates of Factors 1 and 2

We examined 10 privacy-related items in the GPD survey questionnaire in relation to our two factors: GPD items 3, 4, 5, 7, 12, 13, 18, 19, 21, 22, and 28. Respondents who answered Not Sure/Don't Know to any of the items we examined were dropped from the analyses (listwise deletion). There is little point in describing the items or the analyses other than to say that the results are very weak: Correlation and multiple correlation coefficients are invariably low, often less than 0.10. However, the results are real: Results of regression and discriminant function analyses usually produced statistically significant F-ratios. Moreover, the F-ratios are sufficiently large so that statistical significance cannot be attributed to cell sizes. Last, most results are in predicted directions, but not all. To illustrate an unexpected result, GPD item 7 asked respondents what actions they had taken to protect personal information. We predicted that taking action would be more likely among those high than those low on Factor 2

because Factor 2 presumably taps a sense of personal control (see GPD item 2). We found just the opposite. In all, the results of our analyses are not what we expected.

Discussion

The good news, if that is the right term, is that we found a factor that appears to be a conceptual replicate of the Harris-Westin General Concern About Privacy Index. Moreover, just as Harris and Westin (1991) report that respondents' attitudes toward privacy issues were predicted by their index, we also found that our factors often predicted attitudes toward privacy-related issues. However, unlike Harris and Westin's results, our results are, as discussed above, invariably weak (but real).

We cannot explain our results. What factors may be contributing to the results? Table 1 reports the percent of respondents for each privacy type. If privacy types represent levels of concern, we can compare the levels of concern in Table 1 with the percent of respondents at the same levels of concern in the U.S. and Canadian samples. To make this comparison, we followed the Harris-Westin criteria for defining their three levels of concern: High, Moderate, and Low (see above) using the items in Factor 1. However, we developed three "tests" of the comparison.

In the "strict" test, only a response (code or recoded) as 4 (i.e., expressing the highest level of concern) was the basis for which level a respondent was assigned. Respondents who responded with high concern on three or four of the four items in Factor 1 were classified as High Concern; those who responded to two items with high concern were classified as Moderate Concern; those who responded to no items or one item with high concern were classified as Low Concern. The results are in Table 3 in the third and fourth columns.

The second test used our definitions of high concern (see its description on pages 5-7). Our definition is less strict than the Harris-Westin definitions we describe on page 2. The results for the authors' definitions appear in Table 3, columns 4 and 5.

The final test uses a very generous or inclusive definition of high concern. Specifically, if a respondent did not responded with a 3 or 4 rating (recoded so 4 equals the highest level of concern) on all four items, s/he was coded as zero. If the respondent responded with a 3 or 4 rating on any one of the four items, it was classified as 1, and so on, so that those classified as 3 or 4, the Harris-Westin High Concern groups, were respondents with either 3's, 4's, of a mix of these two ratings to three or to four of the items, respectively. Using the Harris-Westin criteria, these were converted into the three levels of concern. The results are in Table 3, columns 6 and 7. In columns 1 and 2 of Table 3 are the 2003 Harris-Westin percentages for a U.S. sample by level of concern (Smith, 2006).

Table 3. The Percentage of Respondents Across Levels of Concern in a Harris-Westin Sample and in GPD Canadian and U.S. Samples.

Approach:	Harris-Westin 2003 results	Strict		Authors'		Generous	
		Canada	U.S.	Canada	U.S.	Canada	U.S.
Low	26	85	80	61	58	32	23
Moderate	64	12	12	28	27	29	29
High	10	3	8	11	15	39	48

Two results are obvious. The U.S. and Canadian samples are relatively similar, although Canadian respondents appear to be less concerned about privacy than U.S. respondents. (We have not to date statistically compared these distributions.) More to the

point, neither the Canadian nor U.S. samples, using any of the three tests, remotely resemble the Harris-Westin distribution.

Moreover, not only are our distributions of respondents across levels of unconcern unlike the Harris-Westin distribution, but also the Not Sure rates in our data and the Harris-Westin data are, for some items, quite different. Harris and Westin (1991) report Not Sure rates for two of their four items: They are (about) 1% and 4%. Our Not Sure percentages, for Factors 1 and 2, are in Table 4.

Table 4. Percent "Not Sure" Responses for Factor 1 and Factor 2 Items for Canadian and U.S. Samples.

	Canadian	U.S.
Factor 1		
GDP Item 6	2.89	4.6
GDP Item 10	1.96	2.05
GDP Item 17	22.87	16.2
GDP Item 23	8.29	7.0
Factor 2		
GDP Item 2	3.19	1.3
GDP Item 11	6.29	4.8

If Harris and Westin's (1991) unreported Not Sure rates are in the same range as the two they reported, then clearly we have higher Not Sure rates for two of our Factor 1 items (GPD items 17 and 23). We also have a somewhat high Not Sure rate for a Factor 2 item (GPD item 11). The differences between the percentages, for each item, in the U.S. and Canadian samples are not large, with item 17 a possible exception; overall, the rank order of the percentages in the two samples are relatively similar. We conclude that national differences do not account for the differences between Harris-Westin and our results.

There are many possible reasons for the non-comparability of our data and the Harris-Westin data in Tables 3 and 4. Perhaps they are the result of item differences,

interview/interviewer differences. and/or sample/sampling differences. We simply do not know. However, we have examined two aspects of our items to help us better understand our situation. First, because of our higher Not Sure rates (Table 4), we examined the distributions of respondents across response alternatives on each of the items in our two factors to see if those distributions were unusual. They are not. Table 5 summarizes results of two tests of normality of, and also reports descriptive statistics for, the six items for both the Canadian and U.S. samples. (Because of how our data files are configured, we could not test for normality for the Canadian and U.S. samples separately.) All the tests of normality are statistically significant ($p<.000$, $df = 1076$).

Table 5. Descriptive Statistics and Tests of Normality for Items in Factors 1 and 2

	Kolmogorov-Smirnov statistic (a, b)	Shapiro-Wilks Test (b)	Mean	Standard deviation
GPD item 2	0.314	0.835	2.76	0.834
GPD item 6	0.263	0.848	2.69	0.886
GPD item 10	0,235	0.874	2.58	0.917
GPD item 11	0.248	0.862	2.83	0.922
GPD item 17	0.282	0.855	2.76	0.816
GPD item 23	0.291	0.850	2.52	0.864

Notes:
a) The Kolmogorov-Smirnov Tests include the Lilliefores Significance Correction.
b) Degrees of freedom for each test of normality was 1076 and all results were significant at beyond the .0001 level.

Second, with respect to the items we selected, we are aware that some researchers are more gifted than others at research. One gift is devising the right measures, that is, devising measures that generate results that support hypotheses. Perhaps our choice of items is less adequate than Harris and Westin's, notwithstanding that our factor 1 suggests we have an index of general concern about privacy. In this regard, Smith (2006) would raise questions about our choice of items. She identified GPD survey items she calls

"loose comparisons" (Smith, 2006, p. 9) to the Harris-Westin item. Table 6 summarizes the GPD items that Smith and we identified as comparable to the Harris-Westin items.

Table 6. A Comparison of Items from the GPD Survey Questionnaire Identified as Comparable to the Four Items in the Harris-Westin Index of General Concern About Privacy.

Harris-Westin	Margulis et al. (2006)	Smith (2006)
Concern about threats to your personal privacy	6, 10, **11**	**11**
Consumers have lost all control….	**2**	**2**
Business organizations seek excessively personal information from consumers.	None	6, 19
Government still invading citizen's privacy	**17**, 23	5, **17**, 18

Although Smith and we agreed on the assignment of three items (they are the items in bold in Table 6), there are plenty of differences. Whatever the reason for these differences, they point to the judgmental problems researchers face when attempting a conceptual replication in lieu of a direct replication. What would have been ideal was a study that included the Harris-Westin General Concern About Privacy Index items, the GPD items that we and Smith propose, independent items that would allow a test for divergent validity, and an appropriate sample, so we could empirically test the structure of general concern about privacy. The results could have provided a useful platform for subsequent investigations.

Finally, there are two factors that may have affected our tests of the relationship between our two factors and their attitudinal correlates in the GPD survey questionnaire.

First, we had removed respondents from all analyses who choose Not Sure as their answer to any of six items we selected. List-wise deletion ensures a clean data set but it also raises a question: Did removing these respondents, who may otherwise have expressed strong opinions on other items, reduce behavioral variance which, in turn, contribute to our weak results? Second, did the large percentage of respondents in the low concern category (in Table 3) also attenuate our results? We don't know.

In conclusion, we repeat what we stated on page 4. We found a general concern factor in our factor analysis. But nothing else turned out as expected and we don't really know why.

Bibliography

Barkema, H.G. and Vermeulen, F. (1997). What differences in the cultural backgrounds of partners are detrimental for international joint ventures? Journal of International Business Studies, 28 (4), 845-864.

Factor Analysis (2005). SPSS Manual (SPSS 14.0, Release 14.0.1). Chicago: SPSS.

Harris, L. and Westin, A. F. (1991). Harris-Equifax Consumer Privacy Survey 1991. Atlanta: Equifax Inc.

Margulis, S. T. (2003). On the status and contribution of Westin's and Altman's theories of privacy. Journal of Social Issues, 59 (2), 411-430.

Pope, Jennifer A. (2003), Breaking Up is Hard to Do: The Dissolution of Business Relationships in the Global Market Place. Unpublished doctoral dissertation, The University of Texas-Pan American.

Smith, E. (2006). Comparing the Globalization of Personal Data Survey on Privacy and Surveillance to Alan Westin's survey results and the privacy dynamic. Unpublished report, Department of Sociology, Queens University, Kingston, Ontario, Canada.

Westin, A. F. (1970). Privacy and freedom. New York: Atheneum.

Westin, A. F. (2003). Social and political dimensions of privacy. Journal of Social Issues, 59 (2), 431-453.