Appendix A: Connecting Model Parameters to Observed Statistics

The following theorem illustrates the relationship between the model parameters (i.e., heterogeneity h and uncertainty u) and the mean and variance of the observed variable R. The proof of the theorem follows directly from Definition 1.

Theorem 1. *With the latent variable model in Equation 4, the population mean μ and variance σ^2 of the observed variable R satisfy*

$$\mu = \frac{(1-u)\cdot(m-1)\cdot\sqrt{1-4h^2}+m+1}{2}, \tag{6}$$

$$\sigma^2 = (m-1)\cdot\left((1-u)\cdot h^2 + u\cdot\frac{m+1}{12}\right). \tag{7}$$

As can be seen from the theorem, the smaller h and u are, the larger the mean response value μ will be. Translating this to the elicitation of privacy concerns, this means that a strong situational strength, which entails a lower heterogeneity h and a smaller uncertainty u, would *inflate* the expressed privacy concerns, consistent with our earlier discussions.

Compared with the observed mean μ, the observed variance σ^2 has an ostensibly subtler relationship with the two model parameters h and u. Specifically, it appears from Theorem 1 that, while an increasing heterogeneity h always increases σ^2, whether a greater uncertainty u increases or decreases σ^2 depends on the heterogeneity level h (whether $h < \sqrt{(m+1)/12}$, to be precise). To this end, it is important to recall from earlier discussions 1) the range limit of $h \in [0, 1/2]$, and 2) the identifiability requirement of $m \geq 4$ (Iannario 2010). With these two restrictions in place, there is always $h \leq 1/2 < \sqrt{5/12} \leq \sqrt{(m+1)/12}$. Thus, a larger uncertainty u always increases the observed variance σ^2 according to Theorem 1. In sum, the relationships between σ^2 and the model parameters h and u are again consistent with the conceptual framework discussed earlier, as increasing either heterogeneity or uncertainty would lead to a larger variability of the observed responses.

Appendix B: Inclusion of Covariates in the Latent Variable Model

In addition to assessing the model parameters at the distribution level, the latent variable model can be elaborated to allow the examination at an individual level, specifically by linking both model parameters h and u to an individual's characteristics modeled as covariates. In this case, the stochastic form of the latent variable model becomes:

$$P(R_i = r) = (1-u_i)\cdot\left[\binom{m-1}{r-1}\cdot\frac{1}{2^{m-1}}\cdot(1+\sqrt{1-4h_i^2})^{m-r}\cdot(1-\sqrt{1-4h_i^2})^{r-1}\right] + \frac{u_i}{m}, \tag{8}$$

for all possible responses $r \in [1, m]$, where R_i is the response provided by the i-th respondent, and h_i and u_i are the model parameters for the respondent. Like in CUB, we link the covariates with the model parameters in the logistic form, the adequacy of which has been verified through extensive experimentation and simulations (Finch and Hernández Finch 2020). That is, with the covariates for h_i and u_i being vectors $\mathbf{v}_i$ and $\mathbf{w}_i$, respectively, we have $P(R_i = r)$ defined as

$$\frac{e^{-\mathbf{w}_i\mathbf{c}_i}}{1+e^{-\mathbf{w}_i\mathbf{c}_i}}\cdot\left[\binom{m-1}{r-1}\cdot\left(\frac{1}{2}+\frac{\sqrt{(1+e^{-\mathbf{v}_i\mathbf{b}_i})^2-4}}{2+2e^{-\mathbf{v}_i\mathbf{b}_i}}\right)^{m-r}\cdot\left(\frac{1}{2}-\frac{\sqrt{(1+e^{-\mathbf{v}_i\mathbf{b}_i})^2-4}}{2+2e^{-\mathbf{v}_i\mathbf{b}_i}}\right)^{r-1}\right] + \frac{1}{m+m\cdot e^{-\mathbf{w}_i\mathbf{c}_i}}, \tag{9}$$

where vectors $\mathbf{b}_i$ and $\mathbf{c}_i$ are the coefficients for covariates in $\mathbf{v}_i$ and $\mathbf{w}_i$, respectively. The interpretations of these coefficients are similar to those in logistic regression.

Appendix C: Parameter Estimations and Fit Statistics

This appendix includes the parameter estimates for the latent variable model as generated from the responses to 1) the 16 items corresponding to the 16 types of private information (Table 1), 2) the 10 perception-related items and 10 expectation-related items corresponding to the 10 types of companies (Table 2). In each table, we reported the sample size, summary statistics of the responses to each item, point estimate and standard error[13] for heterogeneity and uncertainty, and fit statistics including BIC and the absolute fit measure F^2 discussed earlier. In terms of other fit indices, we also tested but did not report the results for AIC and ICOMP, as these results closely resemble those of BIC. Also, we compared the fit of our latent variable model with an alternative model that optimally fits the observed responses to a baseline distribution using the same log-likelihood function. We tested a variety of baseline distributions (e.g., normal, log-normal, etc.) and found similar results, with normal distribution depicted in Tables 1 and 2.

Also included in the appendix are the parameter estimates for our latent variable model from the responses to all 26 items when demographic variables are included as covariates (Table 4). For each parameter, we only listed the survey items with which at least one of the four demographic covariates has a statistically significant coefficient. Since the sample size and summary statistics of responses for these items were reported in Tables 1 and 2, we only reported in this table the point estimate and standard error for each covariate coefficient, and the fit indices (AIC, BIC, and ICOMP) for our model with and without covariates. Note that we did not report the absolute fit measure F^2, as by definition the model with covariates should always yield F^2 higher than or equal to that of the covariate-free model.

The parameter estimates reported in these tables revealed a minor practical issue for plotting heterogeneity in our two-dimensional framework, which is further illustrated in Figure 4. As can be seen from the figure, plotting heterogeneity with the linear scale could lead to a concentration of values on the range [0.4, 0.5], as any ξ between 0.2 and 0.8 would lead to a heterogeneity value in that range. A simple solution is to plot the heterogeneity value with an adjusted logarithmic scale, e.g., $-\log(c - x)$ where x is the raw heterogeneity and c is a constant[14] slightly above the maximum value of x (i.e., 0.5). Figure 4 demonstrates how such a logarithmic scale can better separate survey items with different values of ξ.

[13] A slight complication arises when computing the standard error for a heterogeneity estimate. Recall from the definition of heterogeneity h that two values of ξ with equal distance to but on different sides of 0.5 are "folded" to the same value of heterogeneity h. This transformation breaks the asymptotic normality of the maximum likelihood estimation (without affecting the feasibility of the estimation itself) if we consider h as the underlying parameter. As a result, it is no long appropriate to derive the (asymptotic) standard error according to the Fisher information, as customary in maximum likelihood estimations. To avoid this issue, we report the estimated standard error of ξ as an upper-bound estimate for that of h in the tables. Note that this approximation does not affect any other values reported in the tables.

[14] Note that the exact value of the constant has no substantial impact on the plotted chart.

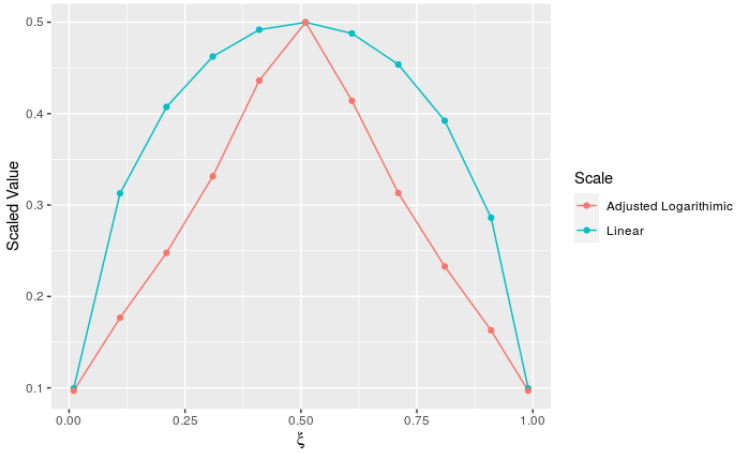

Figure 4 **Context Factor 1: Type of Information Being Collected**

Note. With the adjust logarithmic scale, heterogeneity $h = \sqrt{\xi \cdot (1-\xi)}$ is transformed to $\log(c-h)/\log(c-0.5)$, where c is 0.51 in the chart. With the linear scale, heterogeneity is plotted as is. ξ = CUB model parameter for true attitudes.

Table 1 **Model Estimates for Survey Items Contextualized with Different Types of Private Information**

Item	N	Observation (SD)	Parameter Estimates		Fit Statistics			
			Heterogeneity (SE)	Uncertainty (SE)	BIC	BIC Baseline	F^2	F^2 Baseline
DOB	410	2.04 (1.07)	0.35 (0.06)	0.58 (0.03)***	**1078.75**	1227.50	**.9781**	.7501
SSN	409	1.15 (0.52)	0.11 (0.02)	0.08 (0.00)***	**340.71**	635.24	**.9938**	.4652
browsing	406	2.06 (0.88)	0.47 (0.06)	0.17 (0.02)***	**1010.76**	1059.47	**.9407**	.7329
searching	405	2.11 (0.89)	0.47 (0.06)	0.17 (0.02)***	**1021.15**	1063.96	**.9579**	.7412
emails	405	1.69 (0.91)	0.33 (0.05)	0.30 (0.02)***	**912.87**	1079.85	**.9529**	.6634
SMS	405	1.81 (1.02)	0.29 (0.05)	0.44 (0.02)***	**967.71**	1176.42	**.9908**	.6849
call content	410	1.66 (0.87)	0.34 (0.05)	0.25 (0.02)***	**905.59**	1061.61	**.9482**	.6596
purchasing	410	2.68 (0.80)	0.50 (0.08)	0.00 (0.01)	**989.62**	993.74	**.9457**	.7965
friends	407	2.13 (0.86)	0.48 (0.06)	0.08 (0.02)***	**1013.82**	1044.02	**.9728**	.7458
location	409	1.71 (0.87)	0.37 (0.04)	0.21 (0.02)***	**919.61**	1056.66	**.9958**	.7000
health	409	1.63 (0.87)	0.33 (0.04)	0.23 (0.02)***	**880.02**	1058.03	**.9858**	.6586
politics	408	2.43 (1.03)	0.50 (0.10)	0.61 (0.04)***	**1126.62**	1191.51	**.9929**	.8069
fav media	408	2.79 (0.88)	0.49 (0.07)	0.15 (0.02)***	**1037.44**	1068.77	**.9385**	.8061
call record	407	1.84 (0.93)	0.39 (0.06)	0.30 (0.02)***	**987.57**	1110.22	**.9739**	.7257
religion	409	2.52 (1.08)	0.49 (0.10)	0.82 (0.09)***	**1142.42**	1234.79	**.9814**	.8346
romantic	409	1.92 (1.03)	0.35 (0.05)	0.45 (0.02)***	**1029.82**	1194.62	**.9950**	.7324

Note. See note in Figure 2 for explanations of the item names. N = the number of respondents who provided an answer. Observation = mean response (coded as integer starting with 1 = "very sensitive"). SD = standard deviation. SE = standard error. BIC = Bayesian information criterion (we also tested AIC and ICOMP and observed no qualitative difference). Note that a smaller BIC indicates a better fit. F^2 = absolute fit (Iannario 2012). Baseline = an alternative model fitting the responses with a normal distribution. Bold cells represent the model with the better fit (i.e., smaller BIC or larger F^2). p: *** .001 ** .01 * .05.

Table 2 Model Estimates for Survey Items Contextualized with Different Types of Companies

Item	N	Observation (SD)	Parameter Estimates		Fit Statistics			
			Heterogeneity (SE)	Uncertainty (SE)	BIC	BIC Baseline	F^2	F^2 Baseline
Perception								
landline	353	2.09 (0.94)	0.47 (0.09)	0.18 (0.02)***	**930.32**	970.23	**.8246**	.7724
cellular	370	2.01 (0.92)	0.47 (0.07)	0.12 (0.02)***	**950.41**	996.33	**.8451**	.7617
cable	357	1.97 (0.89)	0.46 (0.07)	0.10 (0.02)***	**896.77**	942.86	**.8872**	.7652
search	368	1.71 (0.81)	0.41 (0.04)	0.06 (0.02)***	**823.19**	898.68	**.9219**	.6953
email	375	1.99 (0.87)	0.47 (0.06)	0.01 (0.02)	**930.08**	970.47	**.8716**	.7721
social media	347	1.57 (0.74)	0.39 (0.03)	0.01 (0.01)	**704.39**	785.38	**.9170**	.6479
video	348	1.60 (0.75)	0.39 (0.03)	0.03 (0.02)*	**720.02**	798.21	**.9415**	.6643
credit card	375	2.27 (0.99)	0.49 (0.10)	0.42 (0.03)***	**1022.76**	1069.25	**.8734**	.8143
online ads	369	1.48 (0.70)	0.35 (0.03)	0.05 (0.01)***	**684.36**	790.27	**.9771**	.6232
ecommerce	376	1.90 (0.86)	0.45 (0.06)	0.09 (0.02)***	**915.34**	968.80	**.9196**	.7566
Expectation								
landline	350	2.81 (1.51)	0.10 (0.04)	0.88 (0.07)***	**1119.67**	1290.87	**.9324**	.8175
cellular	375	2.86 (1.49)	0.10 (0.03)	0.90 (0.10)***	**1204.27**	1375.00	**.9253**	.8237
cable	351	2.50 (1.44)	0.10 (0.04)	0.77 (0.04)***	**1078.33**	1260.85	**.9141**	.7634
search	374	2.06 (1.31)	0.28 (0.05)	0.44 (0.02)***	**1032.64**	1272.72	**.9604**	.7021
email	389	2.56 (1.50)	0.10 (0.04)	0.78 (0.03)***	**1198.73**	1428.84	**.9271**	.7700
social media	338	1.92 (1.24)	0.23 (0.05)	0.41 (0.02)***	**878.71**	1116.25	**.9512**	.6609
video	349	1.75 (1.11)	0.26 (0.04)	0.30 (0.01)***	**837.72**	1075.63	**.9363**	.6314
credit card	373	3.38 (1.41)	0.40 (0.06)	0.64 (0.03)***	**1172.61**	1325.75	**.9247**	.8422
online ads	372	1.65 (1.11)	0.24 (0.04)	0.23 (0.01)***	**804.40**	1143.48	**.9507**	.5850
ecommerce	382	2.57 (1.39)	0.10 (0.04)	0.83 (0.04)***	**1203.35**	1348.34	**.9307**	.7944

Note. Perception: items asking about respondents' expectation of a reasonable time frame for a company to retain their private information. Expectation: items asking about respondents' perceived confidence level on a company keeping their information private. See note in Figure 2 for explanations of the item names. All other notations and terminologies follow Table 1. p: *** .001 ** .01 * .05.

Table 3 Model Estimates for Survey Items Contextualized with Government Surveillance

Item	N	Observation (SD)	Parameter Estimates		Fit Statistics			
			Heterogeneity (SE)	Uncertainty (SE)	BIC	BIC Baseline	F^2	F^2 Baseline
gov:social media	338	2.75 (1.05)	0.45 (0.09)	0.62 (0.04)***	**927.55**	1004.27	**.9888**	.8497
gov:search	394	2.67 (1.01)	0.49 (0.09)	0.52 (0.03)***	**1076.54**	1138.74	**.9966**	.8392
gov:phone	380	2.68 (1.04)	0.47 (0.09)	0.60 (0.04)***	**1043.74**	1118.39	**.9700**	.8573
gov:apps	334	2.80 (1.03)	0.45 (0.08)	0.54 (0.04)***	**904.19**	979.80	**.9828**	.8679
gov:email	393	2.63 (1.07)	0.47 (0.08)	0.72 (0.05)***	**1090.22**	1181.54	**.9721**	.8548

Note. See note in Figure 3(a) for explanations of the item names. All other notations and terminologies follow Table 1. p: *** .001 ** .01 * .05.

Table 4 Model Estimates with Demographics as Covariates

	Covariates				Fit w/ Covariates			Fit w/o Covariates		
Item	Gender	Age	Income	Education	AIC	BIC	ICOMP	AIC	BIC	ICOMP
Uncertainty										
DOB	-0.49 (0.51)	0.07 (0.02)***	0.02 (0.06)	-0.70 (0.34)*	**1061.87**	1102.03	1076.33	1070.71	**1078.75**	**1067.53**
SSN	-0.43 (0.53)	-0.03 (0.02)	0.00 (0.06)	-0.52 (0.25)*	**327.61**	367.75	341.32	332.68	**340.71**	**330.52**
call record	-2.32 (1.44)	0.03 (0.02)	-0.20 (0.10)*	0.03 (0.41)	**977.77**	1017.86	1002.41	979.55	**987.57**	**976.74**
romantic	-0.90 (0.59)	0.02 (0.02)	-0.04 (0.06)	-0.80 (0.32)*	**1003.03**	1043.16	1022.24	1021.79	**1029.82**	**1018.76**
search (E)	-0.14 (0.37)	0.04 (0.01)***	-0.04 (0.04)	-0.37 (0.19)	**1020.06**	1059.30	1034.35	1024.79	**1032.64**	**1022.02**
Heterogeneity										
DOB	-1.62 (1.30)	0.50 (0.39)	-0.05 (0.02)*	-0.01 (0.05)	**1061.87**	1102.03	1076.33	1070.71	**1078.75**	**1067.53**
SMS	-2.43 (0.46)***	0.16 (0.14)	0.03 (0.00)***	-0.01 (0.02)	1038.10	1078.14	1122.09	**959.70**	**967.71**	**956.79**
call content	-2.37 (0.50)***	0.35 (0.16)*	0.02 (0.00)***	-0.03 (0.02)	913.92	954.08	1001.37	**897.55**	**905.59**	**894.88**
purchasing	0.38 (0.32)	-0.11 (0.12)	0.00 (0.00)	-0.03 (0.01)*	988.56	1028.72	1086.98	**981.59**	**989.62**	**979.82**
politics	1.27 (0.42)**	-0.24 (0.15)	-0.00 (0.00)	-0.06 (0.02)**	1126.26	1166.37	1191.04	**1118.60**	**1126.62**	**1115.26**
call record	-2.73 (0.78)***	0.73 (0.29)*	0.01 (0.01)	0.05 (0.03)	**977.77**	1017.86	1002.41	979.55	**987.57**	**976.74**
religion	-3.32 (1.70)	0.15 (0.27)	-0.01 (0.01)	0.25 (0.08)**	**1120.97**	1161.11	1224.85	1134.40	**1142.42**	**1130.44**
cellular (P)	0.30 (0.41)	-0.06 (0.13)	-0.02 (0.00)***	0.01 (0.02)	**937.76**	976.90	1066.84	942.59	**950.41**	**940.32**
search (E)	-1.48 (1.06)	0.32 (0.38)	-0.05 (0.02)**	-0.04 (0.04)	**1020.06**	1059.30	1034.35	1024.79	**1032.64**	**1022.02**
social media (E)	-1.81 (2.03)	0.46 (0.68)	-0.07 (0.03)*	-0.06 (0.08)	874.99	913.22	894.10	**871.06**	**878.71**	**868.45**
video (E)	-0.01 (1.02)	-0.70 (0.45)	-0.01 (0.02)	-0.10 (0.04)*	835.63	874.18	850.61	**830.01**	**837.72**	**827.49**
online ads (E)	-1.65 (0.95)	-0.26 (0.36)	-0.00 (0.01)	-0.08 (0.04)*	803.50	842.69	817.55	**796.57**	**804.40**	**794.04**

Note. (P) = Perception. (E) = Expectation. See notes in Figure 2 for explanations of the item names. Gender = PPGENDER (1: Male, 2: Female). Age = PPAGE (range [18, 86], Mean 48.20, *SD*: 16.59), Income = PPINCIMP (household income, 19 categories, from 1: "Less than $5,000" to 19: "$175,000 or more"); Education = PPEDUCAT (4 categories, from 1: "Less than high school" to 4: "Bachelor's degree or higher"). AIC = Akaike information criterion; BIC = Bayesian information criterion; ICOMP = Information complexity (Bearse et al. 1997). Bold cells represent the model with the better fit (i.e., smaller AIC, BIC, or ICOMP). To assess the (null) hypothesis of a coefficient in $\mathbf{b}_i$ (or $\mathbf{c}_i$) being zero (i.e., no relationship between the corresponding covariate and the model parameter), standard test statistic was constructed from the standard errors generated by the maximum likelihood algorithm for each parameter estimate (cf. Finch and Hernández Finch 2020). For each parameter, we only listed the survey items that yield a non-singular solution with which at least one of the four demographic covariates having a statistically significant coefficient (i.e., $p < .05$). p: *** .001 ** .01 * .05.

Table 5 Model Estimates for Survey Items Contextualized with Facebook and Google

Item	N	Observation (SD)	Parameter Estimates		Fit Statistics			
			Heterogeneity (SE)	Uncertainty (SE)	BIC	BIC Baseline	F^2	F^2 Baseline
Facebook:11	559	2.81 (1.01)	0.46 (0.16)	0.47 (0.06)***	**275.55**	294.38	**.9853**	.8506
Facebook:18	785	3.06 (1.01)	0.38 (0.12)	0.43 (0.05)***	**259.88**	294.62	**.9993**	.8865
Google:11	904	2.58 (1.07)	0.50 (0.20)	0.75 (0.13)***	**284.77**	304.65	**.9603**	.8121
Google:18	1106	2.87 (1.04)	0.42 (0.15)	0.55 (0.07)***	**275.26**	300.05	**.9638**	.8505

Note. See note in Figure 3(c) for explanations of the item names. All other notations and terminologies follow Table 1. p: *** .001 ** .01 * .05.

Appendix D: List of Survey Items

As discussed in the paper, we excluded all non-responses (e.g., refusal to answer to a question, "do not know", "no opinion") from the scales. We also rearranged the options such that smaller values always represent more pronounced privacy concerns.

D.1. Type of Information (Pew Surveys - Wave 1)

Wed now like to know how you feel about a range of information that others might learn about you in daily life. For each kind of information, please indicate how sensitive you consider that information to beeven if some people and organizations already have access to it. (Scale: 1. Very sensitive; 2. Somewhat sensitive; 3. Not too sensitive; 4. Not at all sensitive)	
[Purchasing]	Your basic purchasing habits things like the foods and clothes and stores you prefer.
[Friends]	Who your friends are and what they are like
[Location]	Details of your physical location over a period of time, gathered from the GPS data from your cell phone
[Health]	The state of your health and the medicines you take
[Politics]	Your political views and the candidates you support
[Fav media]	The media you like such as your tastes in music, movies, books, websites, magazines
[Call record]	The numbers you have called or texted from your phone
[Religion]	Your religious and spiritual views
[Romantic]	Your relationship history, including people you have dated or were romantically involved with in the past
[DOB]	Your birth date
[SSN]	Your social security number
[Browsing]	The websites you have visited
[Searching]	Searches you have made using online search engines
[Emails]	The content of your email messages
[SMS]	The content of your text messages
[Call content]	The content of your phone conversations

D.2. Type of Companies (Pew Surveys - Wave 2)

D.2.1. Expectation

Various companies and organizations sometimes need to keep records of your activity on file for their own purposes or because of legal requirements. How long do you think it is reasonable for the following companies or organizations to retain their records or archives of your activity? (Scale: 1. They shouldn't save any information; 2. A few weeks; 3. A few months; 4. A few years; 5. As long as they need to.)	
[Landline]	Your landline telephone company
[Cellular]	Your cellular telephone company
[Cable]	Your cable TV company
[Search]	Your search engine provider(s)
[Email]	Your email provider(s)
[Social media]	The social media sites you use
[Video]	The online video sites you use
[Credit card]	Your credit card companies
[Online ads]	The online advertisers who place ads on the websites you visit
[Ecommerce]	Companies or retailers you do business with

D.2.2. Perception

Still thinking about some of the companies and organizations that maintain records of your activity, how confident are you that these records will remain private and secure? (Scale: 1. Not at all confident; 2. Not too confident; 3. Somewhat confident; 4. Very confident.)	
[Landline]	Your landline telephone company
[Cellular]	Your cellular telephone company
[Cable]	Your cable TV company
[Search]	Your search engine provider(s)
[Email]	Your email provider(s)
[Social media]	The social media sites you use
[Video]	The online video sites you use
[Credit card]	Your credit card companies
[Online ads]	The online advertisers who place ads on the websites you visit
[Ecommerce]	Companies or retailers you do business with

D.3. Government Surveillance (Pew Surveys - Wave 3)

How concerned are you about government monitoring of your _____? (Scale: 1. Very concerned; 2. Somewhat concerned; 3. Not very concerned; 4. Not at all concerned.)	
[Social media]	Activity on social media websites such as Facebook or Twitter.
[Search]	Activity on search engines
[Phone]	Activity on your cell phone
[Apps]	Activity on your mobile apps
[Email]	Email messages

D.4. Facebook and Google (Gallup Surveys - 2011 and 2018)

How concerned are you about _____? (Scale: 1. Very concerned; 2. Somewhat concerned; 3. Not too concerned; 4. Not concerned at all.) *Note: The question was the same in 2011 and 2018.*	
[Facebook]	Invasion of privacy when using Facebook
[Google]	Invasion of privacy when using Google