

# Exploring Respondents' Perceptions of Data Confidentiality and Enhanced Cybersecurity

**Amy Swallow, Robin Kaplan, & Jennifer Edgar**

Bureau of Labor Statistics

Office of Survey Methods Research (OSMR)

FedCASIC 4/12/17



# Background

- Confidentiality can play a role in the respondents response decision ([Mayer, 2002](#))
- Pledges often contain legal language that's unclear to respondents ([Gerber, 2003](#); [Landreth, 2008](#))
- Most government pledges invoke a statute ensuring confidentiality and only statistical uses of data provided
- Despite some variations in trust of the federal statistical system, government surveys are known for high response rates ([Childs, 2015](#))



# Federal 2015 Cybersecurity Enhancement Act

Big law, lots of fine print, complicated IT specifications



Requires the installation of the Department of Homeland Security's EINSTEIN cybersecurity protection system on all Federal civilian information technology systems.

## CIPSEA

vs.

## Cybersecurity Enhancement Act

- Confidential Information Protection and Efficiency Statistical Act
- Only BLS employees, agents and partners can access BLS data
- Data can only be used for statistical purposes

- New cybersecurity technology 'trumps' CIPSEA
- DHS can access BLS information if flagged
- Information can be used to identify and mitigate cybersecurity threats



# Updating Pledge Language

Informed  
Consent

Alarming  
Respondents



# Updating Pledge Language

Informed  
C

Per the Cybersecurity Enhancement Act of 2015, Federal information systems are protected from malicious activities through cybersecurity screening of transmitted data.

nts



# Research Goals

- Understand perceptions, attitudes, and comprehension of current and revised pledge language
- Explore data confidentiality concerns
- Understand potential impact on survey response



# Survey Design

## 1. Baseline questions

- Prior thought about survey confidentiality
- Beliefs about how the government keeps data confidential

## 2. Pledge + Demographics

- Random assignment to pledge type (old or revised)
- Demographic questions

## 3. Memory Questions

- Summarize what they remember from the pledge

## 4. Reread Pledge

- Describe pledge in own words
- Attitude measures



# Survey Design

Question #	Description
1. Baseline	Initial beliefs on how the government keeps data confidential
2. Pledge + Demographics	Random assignment to read the old or revised pledge type and were subsequently asked demographic questions
3. Summarize Pledge	Summarization of recalled information from previously shown pledge
4. Paraphrase Pledge	Pledge was shown again and participants described it in their own words



# Participants

- Mechanical Turk participants randomly assigned to the old (n = 438) or revised pledge (n = 426)
- Demographics statistically equal between groups

Average Age	36 years
Average Annual Income	\$31,346
Average Household Size	2.8 people
College Degree or Higher	63%
Employed Full Time	57%
White	78%

# Completion Rate

	Frequency	Percentage
Partial complete	264	23.4%
Complete	864	76.6%
Total	1128	100%

- No differences in time spent reading each pledge ( $M = 31.4$  seconds,  $p > .05$ )
- No differences found in attrition between pledge conditions

# Coded Questions

- Baseline: (before reading pledge)
  - ▶ "Without looking for any outside information, please tell us how you think the government keeps survey data confidential."
- Summarize Pledge: (after reading pledge)
  - ▶ "In your own words, please summarize the information about confidentiality at the BLS."
- Paraphrase Pledge: (after reading pledge again)
  - ▶ "Now that you see the [confidentiality] text again, please rephrase it using your own words."



# Baseline:

**"Without looking for any outside information, please tell us how you think the government keeps survey data confidential."**

Code Category	Example	Percent
Secure	"I presume the data is stored on secure computer servers"	83.5%
Not secure	"I don't know if they do, because lately it has been in the news that the government is not protecting information"	8.0%
Don't know	"I'm really not sure"	3.7%
Sells information	"They don't sell my information to any companies"	1.7%
No response		1.4%
Other	"ok"	1.7%



# Baseline Response Examples

**"Without looking for any outside information, please tell us how you think the government keeps survey data confidential."**

- "I would hope they are on private servers or the information is kept in a private storing place."
- "I would think they have a computer software system that is heavily password protected to store our data."
- "Probably some archaic, easily hackable way like everybody else who has survey data."
- "It probably doesn't. It probably uses the information to spy on citizens"



# Summarize Pledge

**"In your own words, please summarize the information about confidentiality at the BLS."**

Code Category	Example	Percent
<b>Secure</b>	"They keep the data confidential and do not share it"	57.8%
<b>Don't know</b>	"I don't remember, that was many screens ago"	16.7%
<b>Secure &amp; statistical purposes only</b>	"They keep all data confidential and only use answers for statistical purposes"	15.7%
<b>Statistical purposes only</b>	"The data will only be used for statistical purposes"	5.0%
<b>Cheated</b>	"The Bureau of Labor Statistics, its employees, agents, and partner statistical agencies, will use the information you provide for statistical purposes only..."	2.1%
<b>Sell data</b>	"That my personal information would not be sold off to a third party"	0.5%
<b>Other</b>	"It is a government agency created to gather labor statistics"	2.3%

# Summarize Pledge Examples

**"In your own words, please summarize the information about confidentiality at the BLS."**

- "Personal information will not be released by the BLS without written notification. "
- "Standard government legalese. Data will be protected, Federal Information Security Act, etc., etc. "
- "BLS does the best to promote confidentiality through conscious security and data practices. "
- "I am at a blank now"





# Paraphrase pledge

**"Now that you see the text again, please rephrase it using your own words."**

Code Category	Example	Percent
<b>Secure &amp; statistical purposes only</b>	"They will use your data for statistical purposes only and comply with federal laws to keep your information safe."	48.8%
<b>Secure</b>	"The BLS will keep your data confidential and will not disclose it without my consent"	26.3%
<b>Don't know/no response</b>	"I did not remember anything"	14.3%
<b>Cheated</b>	"The Bureau of Labor Statistics, its employees, agents, and partner statistical agencies, will use the information you provide for statistical purposes only..."	6.9%
<b>Statistical purposes only</b>	"The BLS will only use my personal information for statistical purposes"	2.9%
<b>Other</b>	"Americanism"	0.7%

# Paraphrase Pledge

**"Now that you see the text again, please rephrase it using your own words."**

- "No. It's nonsense"
- "The Bureau of Labor Statistics will keep statistical information confidential and will not disclose information without consent under the law."
- "your information will be used for statistical purposes only and will be protected from malicious activities"
- "The BLS will hold your info confidential to the full extent legally permitted."



# Differences by Pledge Type

- Both pledge conditions had similar understanding across the Baseline, summarize pledge, or paraphrase pledge coded categories (all  $p$ s > .05)
- Both pledge conditions had similarities across other attitude variables (all  $p$ s > .05).

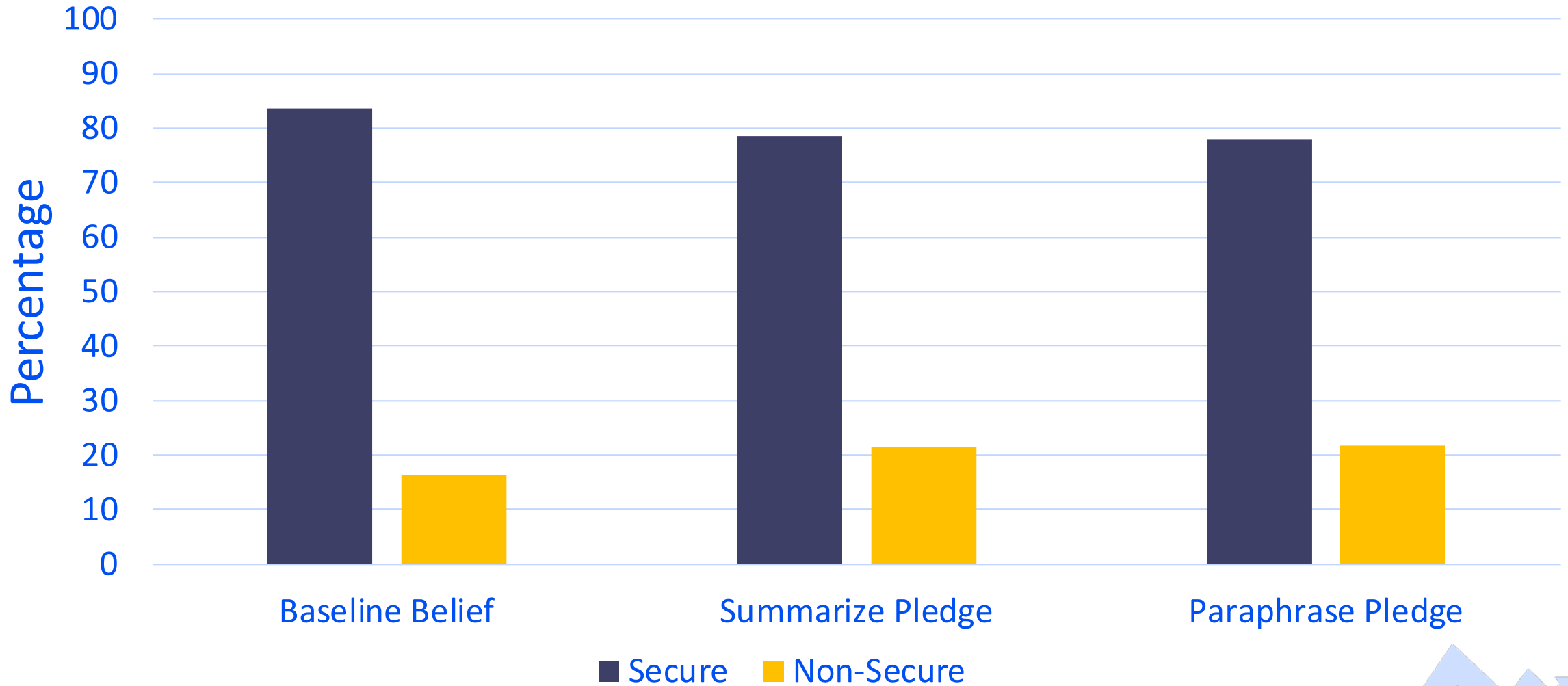
# Results



# Coded Questions

- Baseline: (before reading pledge)
  - ▶ "Without looking for any outside information, please tell us how you think the government keeps survey data confidential."
- Summarize Pledge: (after reading pledge)
  - ▶ "In your own words, please summarize the information about confidentiality at the BLS."
- Paraphrase Pledge: (after reading pledge again)
  - ▶ "Now that you see the [confidentiality] text again, please rephrase it using your own words."

# Distribution of Coded Variables



# Correlations between Coded Variables

	Summarize Pledge	Paraphrase Pledge
Baseline beliefs	0.10*	0.06, n.s.
Summarize pledge	-	0.60*

-Participants with baseline beliefs that their data were secure also later recalled that the pledge said that their data were secure.

-Participants who recalled that their data were secure also stated in their own words that they believed BLS would keep their data secure.

\* $p < .05$

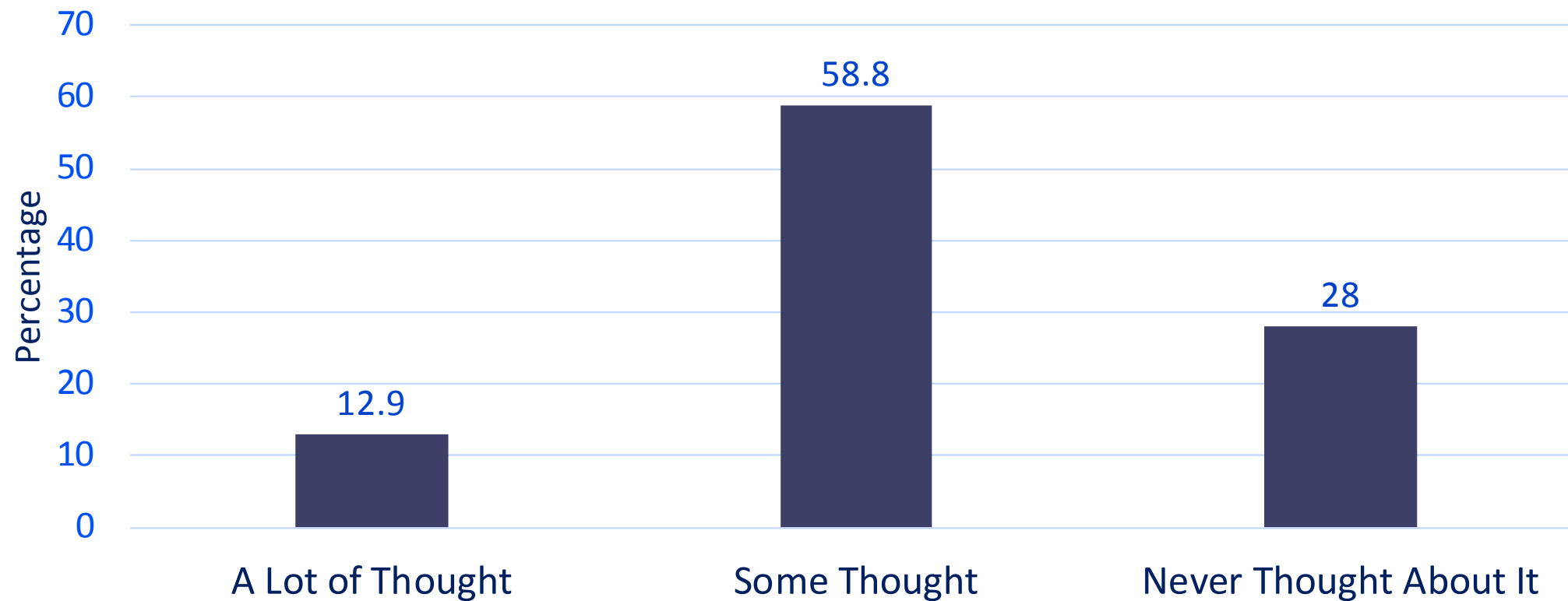


# Baseline Attitudes about Confidentiality





# Thought about Confidentiality



- "Prior to today, how much have you thought about how survey companies (the government or otherwise) protect the information you provide?"

# Correlations between Thought about Confidentiality and Coded Variables

	Baseline beliefs (secure vs. non-secure)	Summarize Pledge (secure vs. non-secure)	Paraphrase Pledge (secure vs. non-secure)
Amount of thought about data confidentiality (none/some vs. a lot)	-.07*	-.05, n.s.	-.07*

\* =  $p < .05$



# Post-Pledge Attitudes about Confidentiality



# Confidentiality Concerns Measure

- "Do you have concerns about how BLS would keep your information confidential?"
  - Yes (21.1%)
  - No (78.8%)



# Correlations between Concerns about BLS Confidentiality and Coded Variables

	Baseline beliefs (secure vs. non-secure)	Summarize pledge (secure vs. non-secure)	Paraphrase pledge (secure vs. non-secure)
Concerns about how BLS keeps data confidential (concerns vs. no concerns)	-0.27**	-0.03, n.s.	-0.12**

\* =  $p < .001$



# Conclusions

- No differences between the old and revised pledge versions across baseline, summarize pledge, and paraphrase pledge variables
- However, differences were found regarding attitudes
  - ▶ Attitudes may play a large role in determining perception of confidentiality
- Attitudes towards confidentiality may impact respondent behavior more than pledge language

# Thank you!



# Supplemental Slides





# Amazon Mechanical Turk

- Website where the general public can participate in research studies for a nominal fee
- Often obtain similar results to surveys using population-based samples
  - ▶ More representative than other convenience samples (i.e. university students)
  - ▶ Served as proxy for BLS household respondents
- These results used as a part of a larger, interagency research project investigating this issue



# Item Non-Response

Item	Old pledge	New pledge
Age	1.7%	.9%
Education	1.1%	0.2%
Employment status	1.1%	1.0%
Occupation	23.0%	21.6%
Job title	22.7%	21.6%
Income	1.3%	0.4%
Household income	1.3%	0.6%
Household size	1.1%	1.5%
Reported Hispanic	1.3%	1.4%
Ethnicity	2.6%	2.6%
Zip code	1.1%	1.7%

# Demographics

- White participants were more likely to recall & say in own words that pledge keeps data secure than non-white participants.
- Younger people more likely to believe data secure at baseline, but older people more likely to recall secure.
- No other findings across any demographic variables were significant.

# Completion Rate and Paradata



# Correlations between Completion Rate and Coded Variables

	Baseline beliefs (secure vs. non-secure)	Summarize pledge (secure vs. non-secure)	Paraphrase Pledge (secure vs. non-secure)
Completion Rate (partial vs. complete)	-0.04, n.s.	0.60*	0.64*

\* =  $p < .001$



# Free Recall of Pledge Words

Recalled Word	% Participants Recalling	% of All Words Recalled
Confidential	2.5	6.4
Information	1.4	3.6
Law	1.2	3.2
Consent	0.9	2.4

"Fine" "International"  
"Penalty" "Anonymous"  
"Secret"

# Contact Information

**Amy Swallow**

Student Intern

BLS/OSMR

[www.bls.gov/OSMR](http://www.bls.gov/OSMR)

202-691-7107

[swallow.amy@bls.gov](mailto:swallow.amy@bls.gov)

