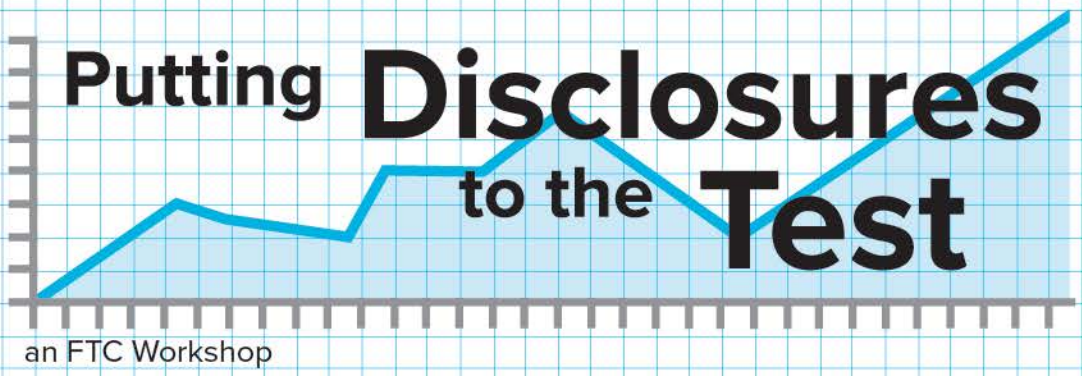


STAFF SUMMARY

November 2016

Table of Contents

1. Introduction

On September 15, 2016, the Federal Trade Commission convened a public workshop, Putting Disclosures to the Test, that examined ways of testing and evaluating the effectiveness of disclosures in communicating a wide range of information that consumers need to make informed decisions in the marketplace. Disclosures may be delivered offline or online through icons, product labels, short text, long text, audio or video messages, interactive tools, and other media. The FTC focuses on disclosures that affect consumer welfare such as disclosures that inform consumers about the risks from using certain products, or disclosures necessary to limit or qualify marketing statements in order to prevent deception. Disclosures may inform consumers about the choices they have and allow them to make accurate comparisons between products and services. For example, disclosures may help consumers exercise control over the way their personal financial information is used,[1] understand the energy consumption of appliances[2] or light bulbs,[3] or understand the costs associated with funeral service[4] options.

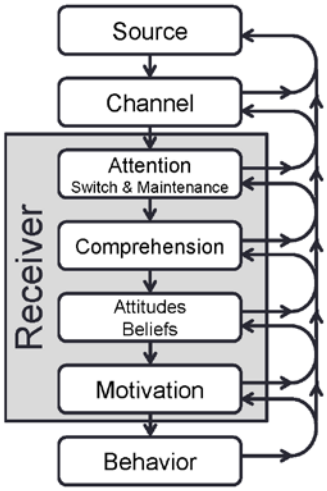
Chairwoman Ramirez opened the workshop[5] by highlighting FTC guidance and enforcement actions related to disclosures and emphasizing the importance of disclosure evaluation. She noted that "ineffective disclosures can overwhelm, confuse, or even distract consumers from making informed choices," and encouraged businesses to "pay attention to ensuring that disclosures provide useful information that can translate into consumer action." She concluded, "We aim to highlight the importance of empirical analysis of disclosures and encourage marketers, businesses, and other organizations to test their own disclosures and learn from researchers."

The FTC has a long history of encouraging meaningful and useful disclosures by conducting or commissioning studies of disclosure effectiveness[6] and bringing enforcement actions against companies that fail to provide adequate disclosures to consumers. The FTC also has issued guidance about making effective digital disclosures[7] and mobile privacy notices[8] and developed standardized disclosure formats in some specific areas. The FTC requires that necessary disclosures be "clear and conspicuous" so that consumers will see or hear the disclosures and understand them. This is so consumers can use the information conveyed to make informed decisions. Disclosures created and deployed without consumer testing may not achieve these objectives.

A number of factors impact the effectiveness of disclosures, including whether they contain the most essential information, and whether consumers notice them, direct their attention towards them, comprehend them, and are able to use that information in their decision-making. Some testing methods are more appropriate than others for evaluating these factors. This workshop brought together experts from a variety of disciplines to discuss disclosure evaluation methods.

This report summarizes the workshop and highlights key takeaways. Slides, videos, transcripts, materials referenced in speaker presentations,[9] and photographs[10] are all available from the workshop website.[11]

2. Cognitive models and evaluation methods



Michael S. Wogalter[12] presented[13] a cognitive framework for assessing disclosure effectiveness known as the Communication-Human Information Processing (C-HIP) Model.[14] This model can be used to describe how humans process disclosures, beginning with when a *source* transmits disclosure information through a *channel* to a human *receiver*. How the receiver processes the information may depend on demographics, training, and other characteristics.

Once a disclosure is transmitted, people need to switch their *attention* to the disclosure and maintain attention long enough to process it. Eye tracking studies can help evaluate whether the receiver notices the disclosure and how long they pay attention to it. Post-study questions can also test whether people remember anything about a disclosure, and thus whether they truly paid attention to it. When the most important information in a long disclosure is highlighted and placed at the top, it increases the likelihood that people will notice and pay attention to it. While people may pay attention to a new disclosure, there is a risk that over time they may become habituated and stop paying attention.

An iterative design and testing process can help improve consumer *comprehension*. While readability metrics are a convenient way to assess comprehensibility, user studies with comprehension tests and interviews are more reliable. Comprehension of text disclosures can often be improved by using simple, active, unambiguous language and an organized structure. Dr. Wogalter mentioned that in some areas standards exist for comprehension testing. For example, he described the ANSI Z535.3 standard, which states that for safety symbols to be acceptable, at least 85% of study participants must correctly understand what the symbol means with no more than 5% critical confusions.

A receiver's *attitudes*, *beliefs*, and *motivation* also play a role in disclosure effectiveness. If a disclosure conflicts with existing beliefs (for example, if people believe something to be lower risk than it actually is), people may be less likely to pay attention and may need salient, persuasive messaging to overcome their erroneous beliefs. In addition, people may not pay attention if they do not believe a disclosure is relevant or that they might learn something new from paying attention. According to Dr. Wogalter, unless people are motivated to apply the information from the disclosure to their decision-making, they will generally take the easiest or most popular course of action.

Finally, Dr. Wogalter noted the importance of assessing the impact of a disclosure on decision-making and *behavior*. This might be measured directly by observing whether people exposed to a disclosure actually change their behavior, or it might be measured indirectly through a survey or by observing some other outcome.

Ilana Westerman[15] discussed a wide variety of disclosure evaluation methods, drawn from many disciplines, and emphasized the importance of using a combination of methods. She recommended that evaluations "ask, observe, and experiment," focusing on using good survey or interview questions to which people can reliably respond, observing behaviors and environmental factors, and systematically

gathering experimental data. Ms. Westerman cautioned against relying on data from improperly executed research and advised taking steps to avoid biasing participants.

Craig Andrews[16] emphasized the need to identify clearly the objectives of an evaluation and metrics before one begins testing. He also discussed the importance of using control groups, choosing an appropriate population and sampling method, and using appropriate analytic techniques. Some of the common testing problems[17] Dr. Andrews highlighted included: assuming that if someone clicked on or saw a disclosure they are aware of and understand it, testing disclosures on populations other than the intended audience, and assuming that the results from a past evaluation will hold under new circumstances. Despite these potential pitfalls, he concluded that, "if you can account for audience characteristics and delivery modes, disclosures can…be effective communication tools and remedies."[18]

The next three panels focused on three aspects of the C-HIP model: attention, comprehension and behavior.

3. Your attention please!

The "Your Attention Please!" panel[19] explored the methods used in research studies that have investigated whether and when consumers pay attention to advertising and privacy disclosures. The studies discussed the impact of text vs. audio modality, wording, presentation, position, and timing of disclosures on attention, and used multiple methods to measure attention.

Nathaniel J. Evans[20] discussed research[21] that evaluated text, audio, and text with audio disclosures intended to inform parents that an "advergame" for kids was a form of advertising. He reported that presenting both audio and text disclosures led to worse advertising recognition than presenting a text disclosure by itself. Dr. Evans suggested that this result might be due to the audio disclosure competing with the sound from the game, thus increasing the user's cognitive load. He recommended that disclosure designers avoid using competing modalities that might increase cognitive load in multimedia game environments.

Mariea Grubbs Hoy[22] presented research that explored whether study participants read a drug risk disclosure in a prescription allergy medicine ad. Eighty percent of participants self-reported reading at least half of the disclosure. However, based on eye tracking data, most had not read the disclosure. Using a retrospective "think aloud" interview, Dr. Hoy determined that participants assumed the drug risk disclosure contained information they already knew or that they did not need to know because the risk was low. She noted the flaws with self-reports about the attention paid to disclosures and urged that they not be relied upon. Dr. Hoy also recommended that disclosures be designed to put information that is most important and likely to be novel up front.

David Hyman[23] discussed a series of large online studies evaluating consumer knowledge and understanding of native advertising[24] and of paid and unpaid content in search engine results pages. Some studies tested words and phrases such as "paid ad" and "sponsored content" to see whether consumers recognized them as indicating that content was paid for by an advertiser. Other studies showed people websites with regular ads and native ads and asked them to identify the paid content.

They compared performance between participants who were shown the ads with no labels, and labels with varying wording, size and color, and position on the page. According to Dr. Hyman, because small changes to ad labels can make a big difference in ad recognition, it is important to conduct experiments to test what is most effective.

Finally, Rebecca Balebako[25] presented two experiments,[26] one in which participants downloaded and used a quiz app on their own smartphones, and one in which participants virtually downloaded and used the same app on their computer screens. Participants in both experiments were divided randomly into groups in which they were presented with a privacy disclosure under different conditions. The disclosure was presented before downloading the app or at varying times while using the app. After completing the quiz, participants answered memory questions about the app, including questions about the privacy notice. Dr. Balebako said that in both experiments, participants presented with the notice prior to downloading the app were significantly less likely to answer questions about it correctly than those shown the notice after the app was installed. She emphasized that timing made a difference in disclosure effectiveness and that the online and field study results were consistent.

In the moderated discussion, panelists discussed the importance of using multiple methods to evaluate disclosures. They noted that budget and time may limit the methods available to some researchers, but said they were generally able to get useful results using relatively low-cost methods such as online studies. Panelists said that eye tracking provided useful data on attention, but that it is currently an expensive and time consuming method (although panelists on a later panel noted that low cost eye trackers are now available). While panelists did not see self-reports as an effective way to assess attention, they found recall and recognition tests to be useful.

4. Comprehension

The "Comprehension" panel focused on methods for evaluating whether people understand the information conveyed in disclosures. The researchers presenting used large online studies or in-person interviews to evaluate comprehension in the context of newspaper articles, nutrition labeling, mortgage disclosures, and privacy notices.

Dan Goldstein[27] presented an approach that helps people understand amounts or other numerical facts by using "perspective" phrases that put the facts in perspective by equating them to quantities with which people may be more familiar. For example, 250 calories might be described as equivalent to 11% of daily calories, 50 minutes of walking, or 31 cups of shredded lettuce. To evaluate the perspectives approaches, researchers presented online study participants with articles with and without the perspectives phrases and measured short-term recall 5 minutes later, long-term recall 3 months later, educated guesses, and error detection. Dr. Goldstein also discussed ways to improve consumer understanding of risks by representing percentages and probabilities with frequencies. For example, Dr. Goldstein said that "1 in 10,000" is much easier for most people to understand than 0.01%. He also discussed research showing that simple visual descriptions of percentages and probabilities can improve understanding. For example, he illustrated how a visual array can help people understand the incidence rate of a disease.

Elizabeth Howlett[28] discussed research[29] on how consumers comprehend front-of-package nutrition labeling. She conducted an online study in which participants were shown food product packages with and without objective and evaluative nutrition icons. Dr. Howlett reported that when participants evaluated a single food item in a noncomparative context, the objective icon was most effective at informing participants about how nutritious the food was. She followed up with a study conducted in a lab that contained grocery store shelves stocked with food products. Dr. Howlett found that when she asked participants to compare, rather than consider in isolation, two similar products, the evaluative icon was most helpful in choosing the healthier food. Thus, she concluded that the context in which people process disclosures makes a difference when evaluating comprehension.

Susan Kleimann[30] discussed the iterative design and evaluation approach[31] she used to develop mortgage disclosures intended to help consumers comprehend information about loans they are considering, compare loans, and choose the best loan for their situation. The disclosures were refined over 18 rounds of qualitative interviews with consumers in English and Spanish. Researchers used a framework known as "Blooms taxonomy"[32] to identify the participants' stages of understanding – knowledge, comprehension, application, analysis, synthesis, and evaluation. Finally, the designs were evaluated in an 858-participant study conducted at 20 locations in which participants were asked to complete comprehension, comparison, and choice tasks using either the existing or proposed disclosures. Dr. Kleimann emphasized: "Comprehension is much more than being able to identify a word or find something in a disclosure. It is really about being able to integrate that information and be able to apply it to yourself so that you understand not merely the technical meaning of something, but the implied meaning."

Joel Reidenberg[33] presented an approach for measuring and comparing vague and ambiguous terms in privacy policies. Dr. Reidenberg and his co-authors examined privacy policies and noted what they described as potentially vague terms, e.g., "may," "generally," "likely," "as needed," and "some." They presented study participants with pairs of sentences from privacy policies that used these terms and asked them which sentence was clearer. They calculated scores designed to correspond with the vagueness of each term, and then applied those scores to all the terms in a policy to rate the vagueness of a policy. Dr. Reidenberg suggested that this approach might be used to develop linguistic guidelines for privacy notices as well as other types of disclosures.

In the discussion, panelists emphasized that to measure comprehension, it is important to ask study participants to perform some analysis and measure their accuracy. They said that in some cases the "right" answer actually depends on the circumstances of each individual, so it is also important to understand the rationale that people use to make a decision.

Panelists also discussed lower-cost approaches to conducting disclosure evaluations, including crowdsourcing platforms and new approaches to eye tracking. Dr. Reidenberg suggested that businesses partner with academic research labs. Dr. Kleimann noted that even small-scale qualitative studies with employees or at a local coffee shop can be useful as long as you keep in mind the narrow demographic profile of your participants in a small study. She said, "My key message would be it is better to test any way than to not test."

Dr. Goldstein explained that because of the expense of surveying large numbers of participants, many past studies have been done with too few participants to understand differences between groups of

people. In addition, he said that previous studies tended to ask every participant a large number of questions due to the cost of recruiting each participant, and that this sometimes resulted in nonsensical answers by the end of a long survey. He pointed out that with the availability of crowdsourcing platforms, such as Amazon Mechanical Turk, where one can hire individuals to complete tasks, large numbers of people can be surveyed at a much lower cost and that it is feasible to ask each participant only a small number of questions. Dr. Howlett added that, using a crowdsourcing platform, she now completes data collection for a study in a single afternoon for a much lower cost than using a marketing research service. Drs. Goldstein and Howlett said journal reviewers have questioned the quality of Mechanical Turk experiments, but there have been studies that have demonstrated that data quality can be just as good and sometimes better when experiments are carried out on Mechanical Turk rather than with traditional methods.

Dr. Goldstein also noted that he has used mouse tracking successfully in online studies as an inexpensive proxy for eye tracking: "This is basically to look at where people are positioning the mouse on the screen. It is a good proxy of where people are looking on the screen, because people read with their mouse more than they think." He added that the price of eye trackers was also dropping and that it is now possible to do eye tracking through a webcam on a laptop rather than investing in $30,000 eye tracking hardware.

Finally, panelists discussed their experiences evaluating the comprehension of icons. Dr. Howlett noted that sometimes people read more into a simple icon than they should. For example, she described a "health halo" effect for a healthy food icon – for example, people may incorrectly assume that a low sodium food is also low fat. Dr. Reidenberg added that in designing privacy icons it has been difficult to figure out how much weight to assign to the many factors that go into evaluating a privacy policy in order to calculate a single grade that would be meaningful for users.

Dr. Howlett also noted the importance of reference points, similar to Dr. Goldstein's perspective phrases. She said consumers do not know how many grams of trans fat or sodium are healthy or how much exercise it takes to work off a certain number of calories.

5. Impact on decision-making and behavior

The "Impact on Decision-Making and Behavior"[34] panelists discussed surveys, online experiments, observational studies, and field experiments that evaluate the impact that disclosures have on consumers' decision-making and behavior. Panelists examined data breach notifications, privacy notices, restaurant and physical report cards, and payday loan disclosures.

Lillian Ablon[35] presented a consumer survey on data breach notifications.[36] The survey was conducted using the "American Life Panel," a nationally representative online panel of American adults. Ms. Ablon noted that while this research method relies on self-reported data, it allows data collection from a representative group with a high response rate. Using this method, she and her colleagues collected data on how often people recalled receiving breach notifications, and what impact people reported these notifications had on their behavior. Ms. Ablon reported that 26% of respondents recalled receiving a breach notification in the preceding 12 months, and more than half of those said they

received more than one notification. She stated that 11% said they stopped doing business with the company that had the breach, 62% said they accepted free credit monitoring services, 51% said they changed their PIN or password, 44% said they already knew about the breach before receiving the notification, and 77% said they were satisfied with the company's response. Ms. Ablon reported that respondents' perceptions of the inconvenience cost of the breach varied from none to over $10,000, with $500 being the median amount for those who reported a non-zero dollar amount.

Idris Adjerid[37] described experiments[38] that examined the impact of non-objective factors on the amount of data people shared after reading a privacy policy. He discussed a Mechanical Turk experiment in which participants were told they would participate in two studies about ethical behavior and in which they were asked sensitive personal questions. Participants saw a "high protection" or "low protection" privacy notice for each study depending on the random condition to which they were assigned. Participants saw the high protection notice for the first study and the low protection notice for the second study or vice versa. Researchers measured the response rate for the sensitive questions and observed that participants were more likely to share personal information when they perceived a relative increase in privacy protection in the second experiment, than when presented with the same high protection notice in the first experiment. Dr. Adjerid said that the results suggest that subtle changes to the framing and presentation of privacy disclosures[39] can have powerful impact. He also said that although this crowdsourced approach may not replicate real-world behaviors and is difficult to use for longitudinal studies, it is a reliable and replicable approach for evaluating the impact of privacy disclosures.

Ginger Jin,[40] the Director of the FTC's Bureau of Economics, analyzed disclosure effectiveness from an economic perspective. She discussed the importance of evaluating not only the effect of disclosure on consumer comprehension and consumer behavior, but also the effect of disclosure on seller behavior. Interaction between both sides of the market can lead to changes in information provision, product price, product quantity, and product quality. Dr. Jin said that the net result can be overall improvement in consumer welfare, improvements for some consumers at the expense of others, or a reduction in consumer welfare. To illustrate, she described studies that examined the impact of mandatory posting of restaurant hygiene grades. Researchers evaluated the impact of mandatory grade posting, reporting that such posting led to improvements in restaurant revenue and sanitary conditions, and reductions in food-borne illness hospitalizations and salmonella infections. However, Dr. Jin noted that in a city where posting the grade was voluntary, many restaurants declined to disclose their grade, even when they had an A grade. In a study involving mandated report cards on physician and hospital cardiac surgery mortality rates, researchers found that physicians reported cherry picking healthier patients and that overall there were higher medical expenditures and worse health outcomes for sicker patients after the report cards were introduced. Dr. Jin noted that "even truthful, quality disclosure can be a double-edged sword." She said that some report cards help consumers make informed choices while others do not, and some encourage sellers to game the system.

Adair Morse[41] discussed her experience engaging with the private sector in testing disclosures aimed at overcoming consumers' cognitive biases in the context of payday loan disclosures.[42] She developed cash envelopes for payday loan customers designed to inform their understanding of either annual percentage rate (APR) or the cost of payday loan fees. Dr. Morse measured each customer's future borrowing after receiving the envelopes. She observed that the envelopes with cost information, but not the envelopes with APR information, reduced future borrowing. Dr. Morse discussed the many challenges of

conducting a controlled field study with customers at multiple locations, as well as the reasons the lender agreed to participate in the study.

Panelists discussed the various methodologies that they used and noted that there were advantages and disadvantages of each. Ms. Ablon noted that the American Life Panel is convenient and provides a representative sample. Dr. Adjerid said that while a field experiment is preferred, crowdsourcing offers the ability to test many more nuanced hypotheses and replicate studies. Dr. Morse added that in a field experiment you usually have only one chance and cannot go back and make changes to your treatments and do another experiment.

6. Case studies

On the next panel, four researchers discussed case studies related to advertising disclosures, drug package labels, and study consent forms. They used a variety of methods including Mechanical Turk studies, lab studies, eye tracking, and mall-intercept studies (participants are people intercepted in a shopping mall and asked if they are willing to participate in a study).

Colin Campbell[43] discussed studies related to native advertising disclosures. In one Mechanical Turk study researchers varied brand familiarity and the position of an ad with a disclosure on a Facebook page. Dr. Campbell said that they found that users had more trouble recognizing ads for unfamiliar brands when the ads appeared in the user's Facebook stream, but not when the ads appeared on the right side of the page. He reported that in a more complex study – one that varied brand familiarity, image professionalism, and disclosure type – they found a threshold effect: ad disclosures appeared to be effective when other ad recognition cues were present but not on their own. Dr. Campbell noted that the "promoted by [brand]" disclosure was more effective than the "advertisement" or "sponsored post" disclosures and that multiple cues contribute to ad recognition, not just the disclosure itself.

Sarah J. Farnsworth[44] discussed the process her firm uses to develop and optimize packaging for over-the-counter medicines, including the drug facts label. She explained that they often initially conduct qualitative studies to understand how consumers think about a product. Once they have an initial design, they conduct label comprehension studies[45] to determine whether consumers can comprehend the information on the label and self-selection studies[46] to determine if consumers can use label information to make accurate decisions about whether or not the product is appropriate for them. Finally, they conduct studies to determine whether consumers can use the product safely in a simulated over-the-counter setting by following label directions on the package. The comprehension studies involve face-to-face interviews with a general population, and are conducted according to FDA-issued guidance.[47] Dr. Farnsworth said that drug facts content with the greatest clinical implication is given the highest weight in the evaluation and required to be understood by a very high threshold of participants, whereas less critical information might have a lower threshold. The testing process generally involves multiple rounds of pilot testing and refinement. It is generally conducted at multiple sites to ensure both demographic and geographic diversity. The FDA typically requests that 20-30% of participants qualify as low literacy. According to Dr. Farnsworth, they test with a general population of consumers rather than only those who have the condition the drug is intended to treat because people may purchase drugs for someone else in the household and may develop a condition in the future. Interviewers read scenario

questions to participants, who reply with open-ended answers. The interviewers then determine which of the pre-coded answer options match the participant's responses. A predetermination is made about which answers are completely correct, and which are acceptable. Dr. Farnsworth thought that this method could be applied to other types of disclosure evaluations.

Manoj Hastak,[48] an FTC consultant, discussed research methods for assessing the efficacy of qualifying information in advertising disclosures. He explained that these studies assess ad communication and ad believability, as well as probe for consumer interpretation of disclosure intent. Dr. Hastak described a mall intercept study of "up to" claims in replacement window ads that showed that rates of communication and believability were similar, regardless of whether the words "up to" were included or whether a disclosure of average results was included in the ad. He also said that a follow-up lab study of "up to" claims in the context of a phone charger ad was consistent with respect to the "up to" qualifier but found that a disclosure of average results had some effect. Dr. Hastak used an eye tracker in the follow-up study and observed that people did look at the disclosure when it was present, and that it did not reduce the amount of time people spent looking at other areas of interest in the ad. He emphasized the value of using multiple measures in disclosure evaluation studies and in replicating previous studies.

Heidi Johnson said the Consumer Financial Protection Bureau[49] is also interested in assessing the efficacy of disclosures, determining what methodologies are most appropriate for evaluating disclosures, and in assessing the effect of disclosure interventions on the marketplace. She said the CFPB uses qualitative methodologies as they develop disclosures and frequently partners with private companies to run field trials. Ms. Johnson said the CFPB also conducts more foundational research to learn generalizable lessons they can apply to future disclosures. She described a lab study focused on the impact of design and context on attention. The study was piggybacked on an unrelated lab study. At the end of that study participants received written information about their privacy rights and payments, which they were asked to sign. The information sheet asked them to initial the form if they were interested in learning about future paid studies. Researchers varied the placement of the information about future studies at the top or the bottom of the form, and whether participants received the form while seated at their workstations or when called up individually to the payment window. Ms. Johnson said that the researchers found only a small effect from form design, but found that people were much more likely to initial the form if they read it at their workstation than if they read it at the payment window. Noting that 100% of participants signed the form, she said had they examined only the signature rate they would have reached a different conclusion about how many people read the form.

The panelists discussed tradeoffs in the methodologies used. Dr. Hastak noted that he had a more diverse population in his mall intercept study, but that the study conducted in a university lab was much less expensive to conduct. He said that he chose a product for the lab study that was likely to be of interest to students. Dr. Campbell said his experience with consumer panels and Mechanical Turk studies is that they produce nearly identical results.

7. The future of disclosures?

The last panel[50] presented studies that evaluate new approaches or new applications of existing approaches to disclosure design and presentations that suggest ways to make disclosures more efficient and effective. The studies focused on mobile app permissions, privacy notices, and medical study informed consent forms.

Serge Egelman[51] described research in which 133 participants' Android phones were equipped to collect data about app usage and the user's browsing, screen locking, and preference data. The researchers collected 176 million events in which apps accessed sensitive data. About five or six times per week for one month participants were prompted with notices informing them about the type of data an app on their phones had just accessed and asking whether they would have allowed this if given the choice. The researchers developed a model of what data would be collected in an "ask-on-first-use" situation where people are prompted for permission only the first time an app tries to access data (the status quo) and how that differed from the preferences participants expressed in their responses to study prompts. According to Dr. Egelman, they determined that 20% of the time participants would have denied subsequent requests and therefore that the ask-on-first-use model produces a 20% error. He said that the researchers then used machine learning to create a model based on users' behavioral data collected for the study. This model correctly predicted whether users would have granted access about 96% of the time, thus reducing the error rate to about 4%. Dr. Egelman said these results raise questions about whether it would be better to ask for permissions to access sensitive data less frequently, using the machine learning model to make permission decisions the rest of the time, or to prompt users more frequently and risk habituating people.

Tamar Krishnamurti[52] discussed research on shortening informed consent disclosures for medical studies based on feedback about what participants find most important. Researchers began with a 17-page consent form for a clinical trial of an asthma treatment. They assigned Mechanical Turk workers who were self-reported asthma patients to read a section of the consent form, select the sentences pertinent to making an informed decision about enrolling in a trial, and rate those sentences on how important they were to their decision-making. They then developed a 5-page consent form based on participants' preferences. Dr. Krishnamurti said that the short form focused on immediate risks and patient experiences in the trial and also included some information deemed essential by the researchers but not requested by participants. In addition, the researchers made a video version that used the short consent form as the script. The researchers conducted a lab study with 76 asthma patients randomly assigned to view the long form, short form, or video. Dr. Krishnamurti reported that participants in all three conditions performed equally well on a knowledge test about the consent form and they all came away with the same perceptions about risks and benefits of the trial. She said that participants who viewed the short form or video reported being more engaged with the form than participants who viewed the long form.

Florian Schaub[53] discussed a study[54] similar to Dr. Krishnamurti's that focused on shortening the privacy policy of a wearable fitness device. In Dr. Schaub's study, researchers first created a succinct version of the policy that presented all of the critical information from the original policy as short bullets. Without viewing the policy, Mechanical Turk workers were quizzed on the policy's terms to determine which terms were already well known. Researchers created medium and short versions of the

policy that removed terms known by either 85% or 70% of participants. Then they tested the effectiveness of the short, medium, and long policies. According to Dr. Schaub, they found that all three policies increased awareness of the company's practices but the shortest policy performed significantly worse than the other policies. In addition, he said that all participants spent about the same amount of time reading each policy, regardless of length. Dr. Schaub also talked about research to develop personalized privacy assistant[55] software for mobile phones that can learn privacy preferences over time.

Panelists noted that some of the approaches to making disclosures more efficient could raise legal and ethical issues. They discussed the question of how to determine the cutoff for what information to present to a consumer and what happens if automated systems make incorrect decisions. Panelists said short or infrequent disclosures could not completely replace full disclosures, but should be viewed as a top layer that would not prevent consumers from going deeper and looking at a full disclosure. They also agreed that auditing or transparency mechanisms were needed, with opportunities for users to provide feedback. Panelists discussed the risk that automated systems might sometimes make mistakes by, for example, consenting to a data collection that a user does not want. Dr. Egelman suggested that users currently have such a poor understanding of privacy policies and choices that even an imperfect automated system might result in decisions that more often match a user's intentions.

Panelists also discussed the need to coordinate multiple versions of policies, and Dr. Schaub suggested that all versions could be derived from a single machine-readable version. Dr. Egelman agreed, and added that machine-readable policies could also help automate the process within a company for keeping privacy policies up to date with actual data practices. Panelists said that machine-readable policies also had an advantage in that they could be presented to the user in a personalized way, and that it would be easier for users to compare multiple policies.

8. Conclusion

Judging from workshop attendance, approximately 225 people in person and 735 remotely via webcast, disclosure evaluation is a topic of broad interest. Researchers presented work from a large number of academic disciplines. Some presenters mentioned that they had not been previously aware of the research being done by researchers in other disciplines that was highly relevant to their own work. There may be benefit in examining disclosures through an interdisciplinary lens to take advantage of differing approaches from fields such as marketing, economics, psychology, computer science, communications, and law.

Panelists suggested various ways to improve disclosure design. They recommended using simple, unambiguous language wherever possible, and an organized structure. Some panelists emphasized that disclosures should be designed with the most important or unexpected information first. Some suggested that user studies can help identify what information is most unexpected or important to users, and that layered disclosures can show essential information on a top layer, with links to more detailed information. Some panelists suggested presenting information that shows people why a disclosure may be relevant to them. Some panelists cautioned that if people see the same disclosure repeatedly they may become habituated and ignore it. Others noted that the timing and context of disclosures can have significant impact on disclosure effectiveness. Some panelists suggested that when presenting numerical

information and risks, it may be useful to put numbers in perspective and express probabilities as frequencies.

Panelists also discussed disclosure evaluation as an iterative activity. Some reported conducting initial research studies to provide insight into what beliefs and knowledge people have prior to receiving a disclosure. Some panelists described iterative design and testing to improve disclosures. It was noted that even inexpensive, small-scale studies can provide insights that can help designers improve disclosures. Some panelists suggested that evaluations may need to be repeated over time as technology, public attitudes, or other circumstances change.

Some panelists recommended using multiple methods to evaluate disclosures, including methods that involve asking participants questions, observing participants, and conducting controlled experiments. They recommended that evaluations and metrics be identified clearly before testing begins, and emphasized the importance of appropriate methods for sampling, experimental design, and data analysis. To evaluate attention, some panelists recommended eye tracking, recall, and recognition tests, and noted that self-reports of attention may not be accurate. When evaluating comprehension, panelists said it is important to evaluate whether participants can apply the information from a disclosure, not just whether they understand all of the words. Some suggested that it is also important to understand the rationale behind the decisions participants make in response to a disclosure.

Some panelists discussed approaches for conducting disclosure evaluations inexpensively. They said that recent developments such as crowd sourcing platforms, mouse tracking, and eye tracking via web cams offer opportunities to conduct larger studies more quickly and less expensively than was previously possible.

Finally, looking towards the future, some panelists discussed using machine-readable disclosures that can be read automatically by software running on each user's device and displayed or acted upon according to each user's context and preferences. They noted legal and ethical issues, but said such an approach could reduce the burden on users to read and manage disclosures and choices, and increase the likelihood that users will pay attention to the information that is most relevant or important to their situation.

Workshop team and staff summary contributors

Lorrie Cranor, Chief Technologist
Joseph Calandrino, Office of Technology Research & Investigation
Laura Hosken, Bureau of Economics
Ryan Mehm, Division of Privacy and Identity Protection
Hampton Newsome, Division of Enforcement
Michael Ostheimer, Division of Advertising Practices
Janis Pappalardo, Bureau of Economics

Endnotes

[1] https://www.consumer.ftc.gov/articles/0222-privacy-choices-your-personal-financial-information
[2] https://www.consumer.ftc.gov/articles/0072-shopping-home-appliances-use-energyguide-label
[3] https://www.consumer.ftc.gov/articles/0164-shopping-light-bulbs
[4] https://www.consumer.ftc.gov/articles/0070-shopping-funeral-services
[5] https://www.ftc.gov/system/files/documents/public_events/950633/ramirez_-_disclosure_workshop_opening_remarks_9-15-16.pdf
[6] https://www.ftc.gov/news-events/blogs/techftc/2016/09/ftc-disclosure-evaluation-research-archives
[7] https://www.ftc.gov/tips-advice/business-center/guidance/com-disclosures-how-make-effective-disclosures-digital
[8] https://www.ftc.gov/sites/default/files/documents/reports/mobile-privacy-disclosures-building-trust-through-transparency-federal-trade-commission-staff-report/130201mobileprivacyreport.pdf
[9] https://www.ftc.gov/system/files/documents/public_events/950633/links_by_panel9-19-16.pdf
[10] https://www.facebook.com/federaltradecommission/photos/?tab=album&album_id=1439278526089000
[11] https://www.ftc.gov/testingdisclosures
[12] http://psychology.chass.ncsu.edu/faculty_staff/mswogalter.php
[13] See part 1 of the workshop video and transcript https://www.ftc.gov/news-events/audio-video/video/putting-disclosures-test-part-1
[14] http://www.safetyhumanfactors.org/wp-content/uploads/2011/12/271Wogalter2006Chap5.pdf
[15] https://iapp.org/about/person/0011a00000DlICIAA3/
[16] http://business.marquette.edu/faculty/directory/craig-andrews
[17] https://www.ftc.gov/system/files/attachments/putting-disclosures-test-materials/pechmann_and_andrews_2011_wiley_international_encyclopedia_of_marketing.pdf
[18] http://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Reports/UCM268069.pdf#page=155
[19] Found at the beginning of part 2 of the video and transcript https://www.ftc.gov/news-events/audio-video/video/putting-disclosures-test-part-2
[20] http://grady.uga.edu/directory/profile/evans
[21] https://www.ftc.gov/system/files/attachments/putting-disclosures-test-materials/parents_presumed_persuasion_knowledge_of_children_s_advergames_the_influence_of_advertising.pdf
[22] https://www.cci.utk.edu/users/mariea-hoy
[23] https://www.law.illinois.edu/faculty/profile/davidhyman
[24] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2816655
[25] http://www.rebeccahunt.com/
[26] http://www.rebeccahunt.com/academic/timing-balebako.pdf
[27] http://dangoldstein.com/
[28] http://marketing.uark.edu/directory.php?id=ehowlett
[29] https://www.ftc.gov/system/files/attachments/putting-disclosures-test-materials/jcr_howlett.pdf
[30] http://www.kleimann.com/teamCore.html
[31] http://www.consumerfinance.gov/know-before-you-owe/compare/
[32] http://bloomstaxonomy.org/Blooms Taxonomy questions.pdf
[33] https://www.fordham.edu/info/23175/joel_reidenberg
[34] Found at the beginning of part 3 of the video and transcript https://www.ftc.gov/news-events/audio-video/video/putting-disclosures-test-part-3
[35] http://www.rand.org/about/people/a/ablon_lillian.html
[36] http://www.rand.org/pubs/research_reports/RR1187.html
[37] http://mendoza.nd.edu/research-and-faculty/directory/idris-adjerid/
[38] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2765097
[39] https://cups.cs.cmu.edu/soups/2013/proceedings/a9_Adjerid.pdf
[40] https://www.ftc.gov/about-ftc/biographies/ginger-zhe-jin
[41] http://facultybio.haas.berkeley.edu/faculty-list/morse-adair/
[42] http://faculty.haas.berkeley.edu/morse/research/papers/BertrandMorsePaydayFieldExpJuly2010.pdf
[43] https://www.kent.edu/business/colin-campbell-phd
[44] http://pegus.com/executiveleadership.html

[45] http://www.accessdata.fda.gov/scripts/cder/training/OTC/topic3/topic3/da_01_03_0170.htm
[46] http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm272122.pdf
[47] http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM143834.pdf
[48] http://www.american.edu/kogod/faculty/mhastak.cfm
[49] http://www.consumerfinance.gov/
[50] Found at the beginning of part 4 of the video and transcript https://www.ftc.gov/news-events/audio-video/video/putting-disclosures-test-part-4
[51] http://www.guanotronic.com/~serge/
[52] https://www.cmu.edu/epp/people/faculty/tamar-krishnamurti.html
[53] https://www.si.umich.edu/people/florian-schaub
[54] https://www.usenix.org/conference/soups2016/technical-sessions/presentation/gluck
[55] https://www.usenix.org/conference/soups2016/technical-sessions/presentation/liu