How Can They Know That? A Study of Factors Affecting the Creepiness of Recommendations

Helma Torkamaan*
University of Duisburg-Essen
Duisburg, Germany
h.torkamaan@acm.org

Catalin-Mihai Barbu*
University of Duisburg-Essen
Duisburg, Germany
catalin.barbu@uni-due.de

Jürgen Ziegler
University of Duisburg-Essen
Duisburg, Germany
juergen.ziegler@uni-due.de

ABSTRACT

Recommender systems (RS) often use implicit user preferences extracted from behavioral and contextual data, in addition to traditional rating-based preference elicitation, to increase the quality and accuracy of personalized recommendations. However, these approaches may harm user experience by causing mixed emotions, such as fear, anxiety, surprise, discomfort, or creepiness. RS should consider users' feelings, expectations, and reactions that result from being shown personalized recommendations. This paper investigates the creepiness of recommendations using an online experiment in three domains: movies, hotels, and health. We define the feeling of creepiness caused by recommendations and find out that it is already known to users of RS. We further find out that the perception of creepiness varies across domains and depends on recommendation features, like causal ambiguity and accuracy. By uncovering possible consequences of creepy recommendations, we also learn that creepiness can have a negative influence on brand and platform attitudes, purchase or consumption intention, user experience, and users' expectations of—and their trust in—RS.

CCS CONCEPTS

• **Information systems** → **Recommender systems**; *Personalization*; • **Human-centered computing** → *User centered design.*

KEYWORDS

Recommender systems; Personalization; Creepiness; Emotion; Trust

ACM Reference Format:

Helma Torkamaan, Catalin-Mihai Barbu, and Jürgen Ziegler. 2019. How Can They Know That? A Study of Factors Affecting the Creepiness of Recommendations. In *Thirteenth ACM Conference on Recommender Systems (RecSys '19), September 16–20, 2019, Copenhagen, Denmark.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3298689.3346982

1 INTRODUCTION

Modern RS algorithms aggregate large amounts of user, item, and contextual data to increase prediction accuracy. However, this level of sophistication is often opaque to users, who might not understand the reasoning behind a recommended item [8]. As a result, people will sometimes feel that a certain recommendation is "creepy" [3].

Creepiness is often the result of over-personalization—especially when users are unaware of its extent and have not stated their preferences explicitly [6]. The perceived lack of transparency and loss of control in the recommendation process can have a detrimental effect on the user's trust in the RS [7]. Thus, exploring the causes and effects of creepy recommendations is a worthwhile endeavor.

2 RELATED WORK

Creepiness is usually defined in the literature as a feeling of uneasiness or perceived emotional harm [16, 17]. It is often caused by the introduction of new technologies, and it is frequently associated with feelings of fear, anxiety, and strangeness [20]. Providing a definitive definition of what constitutes "creepy" has been an elusive task, however. This is mainly due to the term's inherent ambiguity: On the one hand, emotional response is subjective. On the other hand, one's perception of creepiness may evolve over time, e.g., as people reevaluate their own privacy expectations [17].

Most research into what creepiness is and how it appears comes from the field of online behavioral advertising [19]. So far, the attention it has received from the RS community has been rather limited. Several works [3, 14, 18] note the importance of increasing recommendation transparency as a means to make predictions less creepy. Explanations can also trigger feelings of creepiness [6]. In a study on users' perception of computer-generated explanations for advertisements, Eslami et al. [6] found that both vague *and* very specific justifications about why an ad is shown can become creepy. The former might depend on people's tolerance to ambiguity [13], whereas the latter is likely due to users realizing the extent of the tracking. This outcome is especially relevant for RS, where researchers increasingly look at explanations as a means of improving the transparency of their systems [18]. More recently, researchers have also started to argue for more fairness and user control in RS [4], which could also mitigate perceived creepiness.

Langer and König [9] developed a scale for measuring the creepiness of a situation, which can be used as an additional metric for investigating technology-enhanced scenarios. Results show that creepiness correlates positively with privacy concerns and negatively with controllability and transparency [9]. Zhang and Xu [20] developed a theoretical model that considered creepiness a mediator between nudging and privacy attitudes. To our knowledge, there have been no attempts so far to map the feeling of creepiness from a recommendation onto the affective dimensional space.

Based on our literature review, we define the following questions: **RQ1**. How can we describe the creepiness of a recommendation

*Both authors contributed equally to the paper.

RecSys '19, September 16–20, 2019, Copenhagen, Denmark

ACM ISBN 978-1-4503-6243-6/19/09…$15.00
https://doi.org/10.1145/3298689.3346982

using *emotional dimensions*? **RQ2.** Which (a) *user-* and (b) *system characteristics* influence the creepiness of a recommendation? **RQ3.** How does receiving a creepy recommendation influence users' (a) *purchase intention* and *brand attitude*, (b) *expectations about the platform*, and (c) *trust* in the RS?

3 STUDYING CREEPY RECOMMENDATIONS

We conducted an online study using the SoSci Survey platform [11]. The study followed a within-subject design and consisted of four sections: First, we measured user characteristics that could potentially influence the feeling of creepiness. Second, we captured properties of a recommendation that could cause the feeling of creepiness by comparing various domains and conditions. Third, we assessed the implications of a creepy recommendation in terms of users' trust, expectations, purchase intention, and brand attitudes. Finally, we asked basic demographics and seriousness questions. The general setup and tools used in this study are described below.

3.1 User Characteristics

Some user characteristics may influence the perception and feeling of creepiness or discomfort from external entities or stimuli. To consider these characteristics (RQ2a), we used the following constructs and measures: *Social Trust Scale* of the European Social Survey (ESS) [2] to measure the user expectation of trust and fairness (internal reliability Cronbach's $\alpha = 0.86$); The sub-scale *Institution-Based Trust* (IBT) from [12] to record users' trust attitudes toward Internet-based environment in general ($\alpha = 0.81$); The sub-scale *Discomfort with Ambiguity* (DA) from the Multidimensional Attitude Towards Ambiguity Scale (MAAS) [10] to assess users' tolerance of ambiguity, attitudes, and affective reactions toward ambiguous entities or situations ($\alpha = 0.9$); The *Rational-Experiential Inventory* (REI) [5] to measure individual differences in decision-making style ($\alpha = 0.77$).

We measured users' emotional state (RQ1) during the study as a baseline for assessing affect dimensions, i.e. pleasure (P), arousal (A), and dominance (D), using a Self-Assessment Manikin (SAM) [1].

3.2 Recommendation Properties

In order to investigate the properties of personalized recommendations that could cause feelings of discomfort or creepiness (RQ2b), we designed a two-part exploration study. First, we assessed whether participants are familiar with automated online recommendations by explicitly asking about their previous experience. To further ensure that all participants have a basic understanding of RS, we then showed four images depicting recommendations from popular online platforms[1]. The first part of the study concluded by asking participants to think whether they had experienced an online recommendation that had made them feel afraid, anxious, surprised, uncomfortable, or creeped out and to describe the experience in an open-ended answer.

In the second part of the study, we showed participants 9 scenarios (the order was randomized) that described real-life situations at the end of which the user would receive an online recommendation. After reading each scenario, participants were asked to rate, on a 5-point Likert-type scale, the extent to which it would cause them feelings of creepiness if they were to receive such a recommendation. We restricted the scenarios to three domains: movie, hotel, and health recommendations. In choosing the domains, we also considered the participants' likely privacy expectations (i.e. presumably highest in the case of health RS). This allowed us to compare the effect of the domain on users' perception of creepiness. An example scenario from the movie domain is given below:

> "Imagine you searched for a movie on your desktop computer. A few hours later, when you open your streaming app (e.g., Netflix or Amazon Prime), that same movie appears as the first recommendation."

Each domain was featured in three scenarios to assess the following properties of a recommendation: *accuracy*, *causal ambiguity*, and *cross-platform presentation*. Accuracy is defined as the extent to which a recommendation matches the participants' need and preferences exactly. Causal ambiguity appears when users engage in a social activity (e.g., talking about something general) and subsequently receive a recommendation related to the content of that activity. We wanted to understand how participants would react when the reason for receiving a recommendation is unclear. Cross-platform presentation occurs when a recommendation appears on a different medium or platform than what the user interacted with previously. We also captured the emotional aspect of receiving a creepy recommendation using SAM.

3.3 Consequences

To address RQ3, we designed a three-part task that captured how participants' opinions would change after receiving a creepy recommendation. All items were assessed using 5-point Likert scales.

The first task (RQ3a) asked users to rate their impression of: a) the brand, product, or service; b) the platform on which they received the recommendation; c) a future purchase decision of the brand, product, or service; d) a future purchase decision from the platform giving the recommendation; and e) their perception of the usefulness of personalized recommendations. We also asked participants whether their willingness to purchase the recommended item would change if it matched their needs exactly.

The second task addressed RQ3b and contained items that measured users' expectations about the platform in terms of: a) transparency (understanding how their behavioral and personal data was being used for personalization); b) explanation (willingness to read explanations and the reasons behind receiving a specific recommendation); and c) controllability (having the option to decide whether to receive such recommendations in the future). We also asked if the participants' willingness to receive personalized recommendations would change after receiving a creepy recommendation.

The third task measured the change in users' trust in the RS as a consequence of having received a creepy recommendation (RQ3c). We also asked participants whether: a) they considered creepy recommendations to be the result of coincidence; b) they believed that RS are always using their data for personalization; and c) they felt uncomfortable and in need of control.

4 RESULTS AND DISCUSSION

The study was conducted in English, and no special skills were required for participation. Participants were recruited using Amazon

[1] TripAdvisor (hotels), Spotify (music), IMDB (movies), and Amazon (books)

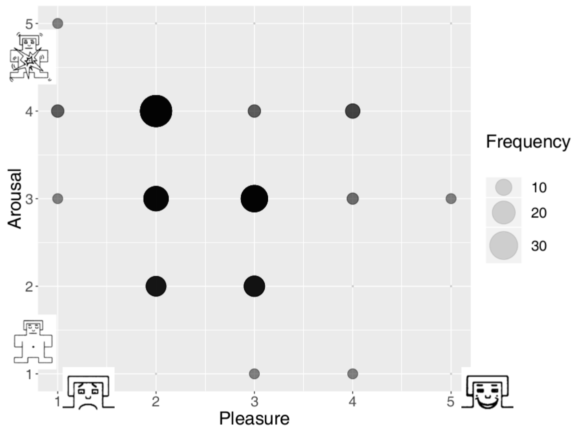

Figure 1: Pleasure and arousal diagram of creepiness with frequency of answers. Axis values correspond to manikins from the SAM scale.

Mechanical Turk and via word-of-mouth. In total, 171 subjects (78 F, 93 M), with an average age of 36.57 ($SD = 11.62$) years, completed the survey in its entirety[2]. Most of them (90%) had used RS before for movies (64.3% of participants), products (63.1%), restaurants (54%), books (50%), music (45%), hotels (42%), or health (17%). Furthermore, the average user had experienced recommendations in several domains ($M = 3.87$, $SD = 2.01$).

Participants described various experiences with RS that had made them feel uncomfortable. Analyzing and categorizing their answers to our open-ended questions revealed that, e.g., the type of product, causal ambiguity, or very high accuracy can trigger feelings of panic, annoyance, or creepiness. Similarly, recommendations for specific product types or services, like those related to a sensitive topic (e.g., recommendations of products for a different age group), can make a person feel uncomfortable or annoyed. Recommendations on delicate topics such as (mental) health were considered particularly creepy.

Recommendations that match users' preferences perfectly can also be perceived as creepy. This could be related to the users' mental model as well as to whether they understand why they are receiving a recommendation. When people are unable to draw a link between the recommendation and their personal preferences—and cannot clearly explain why they received a recommendation—they could experience feelings of creepiness. This might also happen if a user correctly identifies an unwanted modeling of her behavior elsewhere (i.e. on a different platform) that resulted in a personalized recommendation. Several users wrote about situations in which they believed that an arbitrary reason was behind their getting a recommendation and felt creeped out as a result. Participants also mentioned situations where recommendations did not consider their context (e.g., being alone or with others), temporal preference, or mood changes. For some users, recommendations based, for instance, on a (limited) past interaction, a forgotten user history, or an already fulfilled need were also perceived as creepy.

The mixed feeling of *creepiness* as the result of a recommendation is perceived by participants as unpleasant and having a higher level of arousal and a low dominance level (i.e. the feeling of not being in control). Figure 1 shows participants' responses for arousal and pleasure when confronted with a creepy recommendation. The results of our analysis show, on the one hand, how subjective user experiences are: Participants described creepiness with 55 distinct values of P, A, and D on the SAM scale. On the other hand, it indicates, despite the variations, which of the 125 possible triplet values of P, A, and D do not account for the feeling of creepiness from a recommendation. The pleasure dimension of creepiness ($M_P = 2.57, SD_P = 0.93$) is low or very low for almost half of our participants. The arousal level ($M_A = 3.11, SD_A = 0.96$) is high for most of the participants, and the feeling of dominance ($M_D = 2.71, SD_D = 1.21$) is neutral or low when people experience a creepy recommendation. Creepiness as the result of a recommendation appears to fall on the 135° angle of Russell's *circumplex model of affect* [15], adjacent to feelings of frustration, distress, annoyance, nervousness, and fear. This is indicative of a negative user experience; therefore, it seems crucial to consider this effect.

The correlation coefficients between our independent variables (Table 1) are weak to moderate, even for statistically-significant correlations. Interestingly, participants who scored high on social trust also exhibited a lower discomfort of ambiguity. This could mean that they are less likely to consider recommendations creepy. On the other hand, people with a stronger rational style (i.e. REI.NFC) seem to have a higher discomfort with ambiguity. A regression analysis of the independent variables and the SAM dimensions explained only a low percentage of the overall variance. Ultimately, our results for RQ2a are inconclusive and require further analysis.

To investigate RQ2b we performed repeated measures *ANOVA* with an alpha level of 0.05. We considered three domains (i.e. movie, hotel, health) and three recommendation features (i.e. accuracy, causal ambiguity, and cross-platform presentation; see section 3.2). Figure 2 shows the mean creepiness values of the various domains and features. We observed a significant main effect for domain of recommendation, $F(2, 340) = 29.26$, $p < .001$, $\eta_p^2 = .147$. Pairwise

Table 1: Correlation matrix of user characteristics. Variables were assessed using their original Likert scales: ESS (11); DA (7); IBT (7); REI (6). Starred values are significant ($p < .01$).

	M	*SD*	DA	ESS	IBT	REI.FI	REI.NFC
DA	4.50	1.26	1				
ESS	6.68	2.33	-.220**	1			
IBT	5.12	.93	-.053	.441**	1		
REI.FI	4.21	.91	.086	-.021	.320**	1	
REI.NFC	3.54	.66	.373**	.102	.115	.268**	1

Legend: Discomfort with Ambiguity (DA); Social Trust Scale (ESS); Institution-Based Trust (IBT); Faith in Intuition (REI.FI); Need for Cognition (REI.NFC). See section 3.1 for more information.

[2]Out of 324 initial responses, we discarded 153 (47.23%) because either participants did not finish all tasks or the survey was submitted by a problematic (or inattentive) responder. We considered responders problematic if they failed to answer our validity and attention assessment questions—both multiple-choice and open-ended. In addition, we excluded surveys that had been completed in substantially less time than it would realistically take someone to read and answer each question carefully. We, therefore, inferred that those participants had likely not been paying sufficient attention to our survey questions. The remaining cases were deemed valid for inclusion in our analysis.

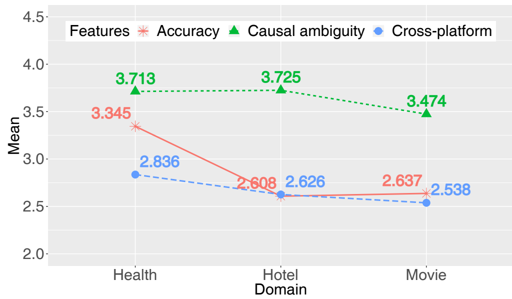

Figure 2: Estimated marginal means of the domains and features. Higher values are more conducive to creepiness.

comparison using Bonferroni correction revealed a significant difference ($p < .001$) between health ($M = 3.30$) and other domains—but not between hotel ($M = 2.99$) and movie ($M = 2.88$) domains. Participants were most sensitive to creepy recommendations in health-related scenarios. The main effect for recommendation features yielded a value of $F(1.6, 277.6) = 98.65$, $p < .001$, $\eta_p^2 = .367$ after Greenhouse-Geisser correction, which denotes a significant difference between features. Posthoc tests using Bonferroni correction further revealed that all features are significantly different from each other ($p < .01$). Causal ambiguity ($M = 3.64$) seems most conducive to feelings of creepiness, followed by accuracy ($M = 2.86$) and platform ($M = 2.67$). The interaction effect was also significant, albeit not very large: $F(4, 426.3) = 10.57$, $p < .001$, $\eta_p^2 = .05$. These results suggest that system characteristics affect the perception of creepiness. Specifically, causal ambiguity in RS leads to a higher level of creepiness. Accuracy and cross-platform presentation are also contributing factors; however, comparing the extent of their influence in various domains requires further investigation.

Figure 3 shows the result of our analysis of RQ3. Different opinions can be observed for the consequences of a creepy recommendation. However, for some of the considered implications, the participants overall exhibit a higher level of agreement with each other. On average, participants' impression of the brand, platform, and their future purchase decision of the brand and platform seem to worsen when they receive a creepy recommendation (RQ3a). However, their impression about the usefulness of a recommendation and their purchase intention if the recommended item matches their preferences remain mostly unchanged. Consequently, we infer that users' impression of the platform, RS, and even brand of the recommended service or product could potentially worsen as a result of a creepy recommendation.

Creepy recommendations also seem to have an effect on user expectations (RQ3b). Participants' willingness to receive personalized recommendations decreases slightly. At the same time, their desire for transparency, explanation, and controllability appears to increase on average. User trust in RS seems to be influenced by creepy recommendations (RQ3c). On average, participants did not believe that these happen by coincidence. At the same time, users had a high level of agreement on whether RS always use their data. As a consequence, they sometimes could form an idea that such systems are spying on them. This may explain why participants, on average, feel the need for more control over RS and why they have lower trust in systems that produce creepy recommendations.

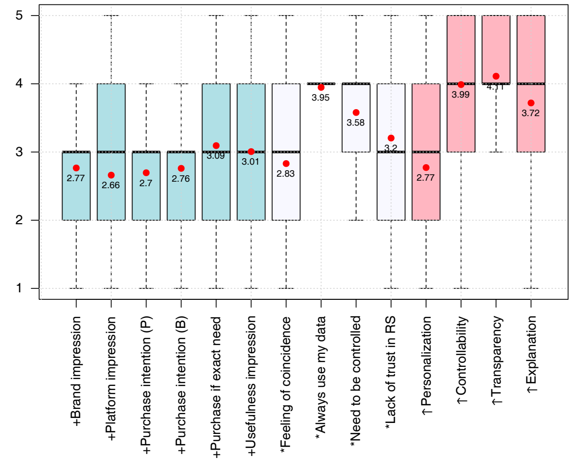

Figure 3: Consequences of a creepy recommendation. (+): Items related to purchase intention and brand attitudes (scale is *worsen–improve*); (∗): trust in RS (*disagree–agree*); (↑): expectations about the platform (*decrease–increase*). Red dots denote mean values.

5 CONCLUSION AND OUTLOOK

This paper provided a comprehensive overview of the creepiness of recommendations by first explaining the user-perceived feeling of creepiness and then by investigating recommendation properties that could contribute to this user perception. It also discussed the consequences of such recommendations. We found that creepy recommendations may cause intense negative feelings—and also that, based on our data, many users have experienced such situations before. Among recommendation properties, perceived creepiness varies across domains. Causal ambiguity seems to contribute more to the feeling of creepiness in comparison with, e.g., cross-platform presentation. Presenting highly accurate recommendations may also cause creepiness if associated with a lack of transparency.

Unwanted personalization is another contributing factor, which can be limited by increasing controllability. Our results suggest that implicit preference elicitation, without transparency and user control over the data collection and personalization of the recommendation, can lead to causal ambiguity and, as a result, to users perceiving recommendations as creepy. These, in turn, may have a negative influence on the perception of the RS or platform, of the recommended brand or service, and on users' trust in RS. It could also increase user expectations of transparency and controllability of the recommendation. Therefore, it seems suitable to try and mitigate creepiness by both keeping users in the loop during preference elicitation and by giving them control over the RS. It is also worthwhile to consider the user's affective state in RS evaluation. In the future, we intend to provide a model indicating the general mechanism (user-system) relevant to the perception of creepiness and to explore additional features and contributing factors.

Acknowledgments: This research is partially funded by the German Research Foundation (DFG) under grant No. GRK 2167, Research Training Group "User-Centred Social Media" and the German Federal Ministry of Education and Research (BMBF) as part of the project PAnalytics (16SV7110).

REFERENCES

[1] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (1994), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

[2] B. Breyer. 2015. Social Trust Scale (ESS). Zusammenstellung sozialwissenschaftlicher Items und Skalen. https://doi.org/10.6102/zis235

[3] Andrea Bunt, Matthew Lount, and Catherine Lauzon. 2012. Are Explanations Always Important? A Study of Deployed, Low-cost Intelligent Interactive Systems. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces (IUI '12)*. ACM, New York, NY, USA, 169–178. https://doi.org/10.1145/2166966.2166996

[4] Michael D. Ekstrand and Martijn C. Willemsen. 2016. Behaviorism is Not Enough: Better Recommendations Through Listening to Users. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. ACM, New York, NY, USA, 221–224. https://doi.org/10.1145/2959100.2959179

[5] Seymour Epstein, Rosemary Pacini, Veronika Denes-Raj, and Harriet Heier. 1996. Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of Personality and Social Psychology* 71, 2 (1996), 390–405. https://doi.org/10.1037/0022-3514.71.2.390

[6] Motahhare Eslami, Sneha R. Krishna Kumaran, Christian Sandvig, and Karrie Karahalios. 2018. Communicating Algorithmic Process in Online Behavioral Advertising. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 432:1–432:13. https://doi.org/10.1145/3173574.3174006

[7] Jason L. Harman, John O'Donovan, Tarek Abdelzaher, and Cleotilde Gonzalez. 2014. Dynamics of Human Trust in Recommender Systems. In *Proceedings of the 8th ACM Conference on Recommender Systems (RecSys '14)*. ACM, New York, NY, USA, 305–308. https://doi.org/10.1145/2645710.2645761

[8] Joseph A. Konstan and John Riedl. 2012. Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction* 22, 1 (Apr 2012), 101–123. https://doi.org/10.1007/s11257-011-9112-x

[9] Markus Langer and Cornelius J. König. 2018. Introducing and Testing the Creepiness of Situation Scale (CRoSS). *Frontiers in Psychology* 9, 2220 (Nov 2018), 1–17. https://doi.org/10.3389/fpsyg.2018.02220

[10] Marco Lauriola, Renato Foschi, Oriana Mosca, and Joshua Weller. 2016. Attitude Toward Ambiguity: Empirically Robust Factors in Self-Report Personality Scales. *Assessment* 23, 3 (2016), 353–373. https://doi.org/10.1177/1073191115577188

[11] Dominik J Leiner. 2019. SoSci Survey (Version 2.5.00-i1142) [Computer software]. Available at http://www.soscisurvey.com/.

[12] D. Harrison McKnight, Vivek Choudhury, and Charles Kacmar. 2002. Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research* 13, 3 (Sept. 2002), 334–359. https://doi.org/10.1287/isre.13.3.334.81

[13] David L. McLain, Efstathios Kefallonitis, and Kimberly Armani. 2015. Ambiguity tolerance in organizations: definitional clarification and perspectives on future research. *Frontiers in Psychology* 6, 344 (April 2015), 1–7. https://doi.org/10.3389/fpsyg.2015.00344

[14] Roberto Pagano, Paolo Cremonesi, Martha Larson, Balázs Hidasi, Domonkos Tikk, Alexandros Karatzoglou, and Massimo Quadrana. 2016. The Contextual Turn: From Context-Aware to Context-Driven Recommender Systems. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. ACM, New York, NY, USA, 249–252. https://doi.org/10.1145/2959100.2959136

[15] James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39, 6 (1980), 1161–1178. https://doi.org/10.1037/h0077714

[16] Omer Tene and Jules Polonetsky. 2014. A Theory of Creepy: Technology, Privacy and Shifting Social Norms. *Yale Journal of Law and Technology* 16, 1, Article 2 (Sep 2014), 44 pages. https://papers.ssrn.com/abstract=2326830

[17] Adam Thierer. 2013. The pursuit of privacy in a world where information control is failing. *Harvard Journal of Law & Public Policy* 36, 2 (March 2013), 409–455. https://ssrn.com/abstract=2234680

[18] Nava Tintarev and Judith Masthoff. 2015. Explaining Recommendations: Design and Evaluation. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, Boston, MA, 353–382. https://doi.org/10.1007/978-1-4899-7637-6_10

[19] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. 2012. Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS '12)*. ACM, New York, NY, USA, Article 4, 15 pages. https://doi.org/10.1145/2335356.2335362

[20] Bo Zhang and Heng Xu. 2016. Privacy Nudges for Mobile Applications: Effects on the Creepiness Emotion and Privacy Attitudes. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 1676–1690. https://doi.org/10.1145/2818048.2820073