Introducing Surprise and Opposition by Design in Recommender Systems

Christine Bauer
Johannes Kepler University Linz
Department of Computational Perception
Linz, Austria
christine.bauer@jku.at

Markus Schedl
Johannes Kepler University Linz
Department of Computational Perception
Linz, Austria
markus.schedl@jku.at

ABSTRACT

There is a long tradition in recommender systems research to evaluate systems using quantitative performance measures on fixed datasets. As a reaction to this narrow accuracy-based focus in research, novel qualities beyond pure accuracy are emphasized in recent research; among them are *surprise* and *opposition*.

This position paper considers that the perception of surprise and/or opposition may be purposely prepared when several recommendations are provided (e.g., in terms of a music playlist) or the user is given the choice between several options.

Altering users' perception and triggering according behavior is well rooted in research on *priming* from psychology and *nudge theory* from the field of economic behavior.

In this position paper, we propose how priming and nudging may be integrated into the design and evaluation of recommender systems to arouse surprise and opposition.

KEYWORDS

recommender system, music, playlist generation, surprise, opposition, priming, nudging, perception, serial recommendation

ACM Reference format:
Christine Bauer and Markus Schedl. 2017. Introducing Surprise and Opposition by Design in Recommender Systems. In *Proceedings of UMAP'17 Adjunct, July 09-12, 2017, Bratislava, Slovakia, ,* 4 pages.
DOI: 10.1145/3099023.3099099

1 INTRODUCTION

For the longest time, recommender systems have been (and are still most commonly) evaluated using quantitative performance measures on fixed datasets. Such measures were originally developed for machine learning and information retrieval tasks and include *error measures* for rating prediction, like root mean squared error and mean average error, *rank-based measures* like Spearman's or Kendall's rank-order correlation coefficients, and *effectiveness measures* such as (mean average) precision, recall, F-measures, or normalized discounted cumulative gain (nDCG).

As a reaction to this narrow accuracy-based focus in research, novel recommender system qualities beyond pure accuracy are emphasized in recent research; among them are qualities such as *diversity* [29, 33], *novelty* [5, 6, 29], *serendipity* [24, 34], *discovery* [4], and *unexpectedness* [1].

In this position paper, we focus on *surprise* and *opposition* and propose to integrate these aspects into design and evaluation of recommender systems. Thereby we will use examples from the music domain. Still, the ideas will mostly also transfer to other domains where recommender systems suggest several items in a row and where the items recommended are intended to entertain the user (e.g., videos or jokes).

Surprise is similar to concepts such as serendipity or unexpectedness. In essence, surprise relates to integrating variations of known elements in unknown ways and/or unpredictable system response and behavior [12].

Opposition is "an extreme form of variation or dissimilarity" [12]. The perception of opposition is very subjective and depends on the context. While, for instance, some people may perceive jazz as "the opposite" of heavy metal, others may find common ground of these two genres [12]. This relates to the question of how to measure dissimilarity. In recommender systems, effecting top-N recommendations based on some measure of similarity is a common approach. While the definition of an accurate function that universally quantifies similarity is already a hard, if not infeasible, task [13, 18, 20, 28], the task to be solved to account for opposition, i.e. quantifying and determining *dis*similar items is even harder [22] and influenced by subjective factors.

Typically, existing recommender systems research starts from the point as to that the recommender system has to identify and suggest items that both, match the user's preferences and/or interests and still are outside the user's typical comfort zone. For instance, most users typically consume mainstream music (in most cases in terms of top artists), while music items from the long tail are in comparison rarely listened to. Recommending music items farther away from mainstream that a user's typical mainstream listening profile would be outside that user's typical comfort zone [23, 30].

Among the main difficulties of this task is to find the right balance between inside and outside comfort zone, such that the user perceives the recommendations as a surprise or opposition and does not perceive the suggested item a result of a poor quality recommender system (in other words, thinking that the recommender system does not fulfill its job in finding a recommendation that fits the user).

However, in domains where a recommender system typically suggests a *series* of recommendations (i.e., continuous or serial recommendations) as for instance in the music domain with playlists, there is another option: A user's perception can be primed and triggered, by leveraging the connections or transitions between

UMAP'17 Adjunct, July 09-12, 2017, Bratislava, Slovakia
 978-1-4503-5067-9/17/07.
DOI: 10.1145/3099023.3099099

consecutively recommended tracks. Consequently, a recommender system can be designed in a way to purposely create qualities such as surprise or opposition. We could call it "Surprise by Design" or respectively "Opposition by Design".

How could this work? For instance in a playlist, after four smooth jazz ballads an up-tempo Bebop song may perceived surprising in a stronger way than after four songs with each increasingly higher tempo. Likewise, when a set of movies starring actor Arnold Schwarzenegger is suggested, among action movies such as the "The Terminator" series, "The Predator", or "Collateral Damage", the tongue-in-cheek comedy "Twins" may be perceived as surprise, compared to suggesting the latter among other satiric or comedy movies such as "Kindergarten Cop" and "Last Action Hero". Similarly, after some satiric poems, a poem on sad love may be perceived emotionally more evoking compared to being presented among other sad poems, since the theme change comes as a surprise. In other words, the perception of surprise may be triggered if it is well prepared.

Varying perception and according purposeful design of systems is well rooted in research on priming from psychology and nudging from the field of economic behavior.

In the next section, we outline the conceptual foundation of our work of priming and nudging. In Section 3, we describe our proposal, before we finally conclude our paper with an outlook to interesting research in the field.

2 CONCEPTUAL FOUNDATIONS

2.1 Priming

Priming (e.g., [2, 10, 15, 25, 31]) is a concept from psychology. It refers to an implicit, non-conscious memory effect in which the exposure to a stimulus influences the response to another stimulus. In other words, the processing of a target stimulus is purposely altered by presenting another specific stimulus beforehand.

A major part of research on priming is based on textual tasks. A typical example is the word-stem completion task (e.g., [19]). Here, participants are given a list of words to study (the stimulus). Then, they are asked to complete word stems (e.g., the first three letters of a word) with the first word that comes to their mind. A priming effect is observed when participants complete stems with words from the list they had to study beforehand than novel words not in the study list.

2.2 Nudging

A similar concept to priming is nudge theory from behavioral economics [8, 27, 32]. This theory proposes positive reinforcement and indirect suggestions to try to achieve non-forced compliance to influence among others the decision making of individuals or groups.

A nudge is thereby any aspect that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives [8]. For instance, for promoting healthy diet, putting fruits at eye level counts as nudge, while banning junk food does not. Similarly, informing clients that a doctor's appointment is scheduled for the next day is nudge. So are the default settings on computers or smartphones [26]. For an overview of literature on nudging in digital environments, see [17].

3 PROPOSAL FOR SURPRISE AND OPPOSITION BY DESIGN

The basic idea we propose here is to use priming and nudging to alter users' perception of the music they are recommended and/or change users' music consumption behavior. While the first is based on priming, the second deploys nudge theory.

3.1 Priming for Surprise and Opposition

When creating a music playlist, priming may be deployed to arouse surprise and/or opposition. Depending on what has been played first, an upcoming song may be perceived surprising because it was not expected to be the next song (or be in the playlist at all) or the song is quite the opposite from what was expected, so arousing opposition. Very trivial examples are: an up-tempo song following a sequence of slow songs; a sequence of songs from one genre and, suddenly, some different genre; a sequence of songs of the same artist, then another artist; etc.

However, creating "good" surprise or opposition is not as trivial as those examples may indicate. For instance, surprise is not always a positively connoted surprise. If you expect another smooth jazz song, a death metal song may come as a surprise, but it may not satisfy the interest of the user. However, some users may enjoy this sequence combination, while others may not enjoy it. While some will perceive it surprising, others will not, for instance, if the stimulus was not sufficiently strong for the respective user. Some users may learn quickly that after, for instance, some slow songs, there will be an up-tempo song surprise, and after a while they may not perceive this sequence surprising anymore, because they even expect it (and would be even surprised if this learned system behavior would not happen once again). For instance, while listening to music while engaged in office work, surprises may not be perceived as strong as compared to when listening to music attentively as the primary task. Furthermore, a music surprise while engaged in work may be perceived as positive surprise, while the same song sequence (including surprise) could be perceived disturbing while jogging (for instance, from empowering songs switching to spherical music). Similarly, in a playlist labeled as "reggae playlist", a song by Frank Sinatra would be surprising but could be annoying to the user as the surprise does not fit the playlist label (so in this case only reggae songs would be expected, and probably also accepted). Also a Christmas song in July would be surprising, but not in a positive way.[1]

Overall, we may summarize that priming for surprise has to account for individual, situational, temporal, cultural, and, in general, contextual differences.

3.2 Nudging for Surprise and Opposition

While priming targets users' perception of music items in a given playlist, nudging aims at altering a user's music consumption behavior for provided suggestions. Thereby, we consider recommending a list of music items from which the user may choose to be the unit to study; similarly, we also consider a pre-assembled sequence of songs (playlist) a unit. Basically, the sequence is treated as a choice

[1]This is only true for users from cultures where Christmas songs are associated with a certain time period in the year.

of songs with the first song in the playlist as the "default' option, which will happen (be played) if the user does nothing.

For instance, experiments and observational studies have shown that making an option a default increases the probability that it is chosen. This is referred to as the default effect [7, 11].

Similarly to priming, we also have to consider differences between individuals, cultures, etc. in nudging. For instance, the default option may be perceived as the "mainstream" option. There will be differences to be considered in personalized music recommendation, if the user wants to happily follow the mainstream or wants to avoid it (in some subcultures, following the mainstream may be perceived "uncool"). In analogy, individuals may tend to follow the crowd (the default option) or not as primary influencing factor, and only in second place may consider the actual content of the option.

Furthermore, frequently the importance of fostering cultural diversity is voiced. A nudge may be used to promote diversity in music consumption based on what is offered as the default option in a list of recommendations or a playlist. Whether a user "tries out" previously unknown music items or ignores them depends largely on the way the items are presented; for instance, whether a previously unknown item is provided as default or somewhere in the last (almost hidden) options. Thereby, the new item may be considered the "opposite" in terms of being new compared to other items. A previously unknown item may also come as a surprise for a user among, for instance, well-known "epic rock anthems" (e.g., "We will rock you" by Queen or "Smoke on the water" by AC/DC). Again we see the importance of the context: For instance, if the user has never heard the rock anthem before (despite the fact that the anthem is generally popular), then the contrast between popular song and new song (i.e., the opposition) may not be perceived at all.

Based on the assumption that users tend to follow the default option, this means for recommender systems, that a surprising item should be among the sequence of default options to be perceived as surprise. Similarly, by designing the default options in a particular sequence, the behavior that many users will follow this path through default options is very likely. Accordingly, such a path may purposely be designed to lead to a surprise.

4 FUTURE WORK

In this position paper, we proposed to exploit priming and nudging effects in the design of recommender systems in order to purposely create perceptions of surprise and/or opposition. We exemplified the idea based on music recommender systems, where several music recommendations are suggested to choose from or in form of playlists (i.e., a sequence of suggestions).

Although priming as well as nudging are rooted in well-established theories, their transfer to and application in music recommender systems open up a new research area, which requires a holistic approach, integrating knowledge from various disciplines (e.g., computer science, psychology, economics, law) and perspectives (e.g., user, platform provider, music creators, artists, labels).

Fundamental research questions still have to be addressed, such as how to prime music perceptions "generally" in recommendations and playlists and how to prime in specific cases, in terms of individual, situational, temporal, and contextual differences. As priming for surprise and/or opposition is inherently contextual, specific user studies for the various settings are necessary [9, 14, 16]. For instance, after some sad songs associated with rainy weather, the surprise effect of suggesting the song "Aquarius (Let the Sunshine in)" from the musical "Hair" may be perceived stronger when driving on a lonely street in pouring rain compared to receiving the same recommendations when listening to music while working late night in the office (e.g., the weather may support associations with the songs' lyrics, while working there is probably less attention paid to background music, the listener's lack in language skills may decrease the likelihood of noticing surprises associated with the lyrics). Accordingly, laboratory studies have to be carried out carefully, controlling for confounding variables inherent in the contextual setting of the experiment. As generalizing from a specific contextual setting to a wider context is difficult or even unfeasible, reporting the context of a study as detailed as possible is fundamental for allowing other researchers to build on findings, compare settings and results, modeling scenarios, integrate those in recommendation algorithms, etc. Building on findings from such studies, algorithms have to be developed that can capture and exploit those findings in order to result in successful recommender systems.

A significant research task in the field of user modeling relates to modeling individual "priming profiles" and integrate them in user models. The same applies to individual "nudging profiles". As both, priming and nudging, are deeply contextual, advances in contextual modeling will also influence priming and nudging effects in recommender systems.

Overall, we believe it is important to integrate into the research strategy a number of small-scale laboratory studies tailored to the respective specific context (as described above) [14]; only then it will be possible to advance the research field with field studies in real-world recommendation settings. Exploiting large-scale open datasets (e.g., the Million song dataset [3] or LFM-1b [21]), which is well-established in recommender systems research, is also a fruitful approach to investigate priming and nudging topics.

Eventually, we emphasize that research in the field of serial recommendation, such as music playlist continuation, needs to consider longer sequences since, for instance, the prediction of solely the respective next item, which is a common approach, ignores existing priming and nudging effects.

5 ACKNOWLEDGMENTS

This research is supported by the Austrian Science Fund (FWF): P25655 and V579.

REFERENCES

[1] Panagiotis Adamopoulos and Alexander Tuzhilin. 2014. On Unexpectedness in Recommender Systems: Or How to Better Expect the Unexpected. *ACM Transactions on Intelligent Systems and Technology* 5, 4, Article 54 (Dec. 2014), 32 pages. https://doi.org/10.1145/2559952

[2] John A. Bargh and Tanya L. Chartrand. 2000. The mind in the middle. *Handbook of research methods in social and personality psychology* (2000), 253–285.

[3] Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. 2011. The Million Song Dataset. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*. Miami, FL, USA.

[4] Dmitry Bogdanov. 2013. *From music similarity to music recommendation: Computational approaches based in audio features and metadata.* PhD. Universitat

Pompeu Fabra, Barcelona, Spain.

[5] Òscar Celma and Perfecto Herrera. 2008. A New Approach to Evaluating Novel Recommendations. In *Proceedings of the 2nd ACM Conference on Recommender Systems (RecSys)*. Lausanne, Switzerland, 8.

[6] Charles L.A. Clarke, Maheedhar Kolla, Gordon V. Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. 2008. Novelty and Diversity in Information Retrieval Evaluation. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '08)*. ACM, New York, NY, USA, 659–666. https://doi.org/10.1145/1390334.1390446

[7] Isaac Dinner, Eric J Johnson, Daniel G Goldstein, and Kaiya Liu. 2011. Partitioning default effects: why people choose not to choose. *Journal of Experimental Psychology: Applied* 17, 4 (2011).

[8] Daniel G. Goldstein, Eric J. Johnson, Andreas Herrmann, and Mark Heitmann. 2008. Nudge your customers toward better choices. *Harvard Business Review* 86, 12 (2008), 99–105.

[9] Daniel M. Hausman and Brynn Welch. 2010. Debate: To nudge or not to nudge. *Journal of Political Philosophy* 18, 1 (2010), 123–136.

[10] E. Tory Higgins, William S. Rholes, and Carl R. Jones. 1977. Category accessibility and impression formation. *Journal of Experimental Social Psychology* 13, 2 (1977), 141–154.

[11] Eric J. Johnson and Daniel Goldstein. 2003. Medicine. Do defaults save lives? *Science (New York, NY)* 302, 5649 (2003), 1338–1339.

[12] Peter Knees, Kristina Andersen, Alan Said, and Marko Tkalcic. 2016. Workshop on Surprise, Opposition, and Obstruction in Adaptive and Personalized Systems (SOAP). In *UMAP 2016 Extended Proceedings – Workshop on Surprise, Opposition, and Obstruction in Adaptive and Personalized Systems (SOAP)* (July 16, 2016). ACM, New York, NY, USA.

[13] Peter Knees and Markus Schedl. 2013. A Survey of Music Similarity and Recommendation from Music Context Data. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)* 10, 1 (2013).

[14] Mark Kosters and Jeroen Van der Heijden. 2015. From mechanism to virtue: Evaluating Nudge theory. *Evaluation* 21, 3 (2015), 276–291.

[15] Karl Spencer Lashley. 1951. The problem of serial order in behavior. In *Cerebral mechanisms in behavior*. 112–136.

[16] Susan Michie and Robert West. 2013. Behaviour change theory and evidence: a presentation to Government. *Health Psychology Review* 7, 1 (2013), 1–22.

[17] Tobias Mirsch, Christiane Lehrer, and Reinhard Jung. 2017. Digital Nudging: Altering User Behavior in Digital Environments. In *Proceedings of13th International Conference on Wirtschaftsinformatik (WI '17)*. AIS.

[18] Alberto Novello, Martin F. McKinney, and Armin Kohlrausch. 2006. Perceptual Evaluation of Music Similarity. In *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR)*. Victoria, BC, Canada.

[19] Roger Ratcliff and Gail McKoon. 1996. Bias effects in implicit memory tasks. *Journal of Experimental Psychology: General* 125, 4 (1996), 403.

[20] Markus Schedl. 2012. #nowplaying Madonna: A Large-Scale Evaluation on Estimating Similarities Between Music Artists and Between Movies from Microblogs. *Information Retrieval* 15 (June 2012), 183–217. Issue 3–4.

[21] Markus Schedl. 2016. The LFM-1B Dataset for Music Retrieval and Recommendation. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval (ICMR '16)*. ACM, New York, NY, USA, 103–110. https://doi.org/10.1145/2911996.2912004

[22] Markus Schedl, Arthur Flexer, and Julián Urbano. 2013. The Neglected User in Music Information Retrieval Research. *Journal of Intelligent Information Systems* (July 2013).

[23] Markus Schedl and David Hauger. 2015. Tailoring Music Recommendations to Users by Considering Diversity, Mainstreaminess, and Novelty. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '15)*. ACM, New York, NY, USA, 947–950. https://doi.org/10.1145/2766462.2767763

[24] Markus Schedl, David Hauger, and Dominik Schnitzer. 2012. A Model for Serendipitous Music Retrieval. In *Proceedings of the 2nd Workshop on Context-awareness in Retrieval and Recommendation (CaRR)*. Lisbon, Portugal.

[25] Lowell H. Storms. 1958. Apparent backward association: A situational effect. *Journal of Experimental Psychology* 55, 4 (1958), 390.

[26] Cass R. Sunstein. 2014. Nudging: A very short guide. *Journal of Consumer Policy* 37, 4 (2014), 583–588.

[27] Richard H. Thaler and Cass R. Sunstein. 2008. Nudge: Improving Decisions About Health, Wealth, and Happiness. (2008).

[28] Julián Urbano. 2013. *Evaluation in Audio Music Similarity*. Ph.D. Dissertation. University Carlos III of Madrid.

[29] Saúl Vargas and Pablo Castells. 2011. Rank and Relevance in Novelty and Diversity Metrics for Recommender Systems. In *Proceedings of the 5th ACM Conference on Recommender Systems (RecSys)*. Chicago, IL, USA, 8.

[30] Gabriel Vigliensoni and Ichiro Fujinaga. 2016. Automatic music recommendation systems: do demographic, profiling, and contextual features improve their performance?. In *Proceedings of the 17th International Society for Music Information Retrieval Conference* (August 7-11, 2016) *(ISMIR '16)*. 94–100.

[31] Markus Weinmann, Christoph Schneider, and Jan vom Brocke. 2016. Digital Nudging. *Business & Information Systems Engineering* 58, 6 (2016), 433–436. https://doi.org/10.1007/s12599-016-0453-1

[32] Peter Wright. 2002. Marketplace Metacognition and Social Intelligence. *Journal of Consumer Research* 28, 4 (2002), 677. https://doi.org/10.1086/338210

[33] Mi Zhang and Neil Hurley. 2008. Avoiding Monotony: Improving the Diversity of Recommendation Lists. In *Proceedings of the 2nd ACM Conference on Recommender Systems (RecSys)*. Lausanne, Switzerland, 8.

[34] Yuan Cao Zhang, Diarmuid Ó. Séaghdha, Daniele Quercia, and Tamas Jambor. 2012. Auralist: Introducing Serendipity into Music Recommendation. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining (WSDM '12)*. ACM, New York, NY, USA, 13–22. https://doi.org/10.1145/2124295.2124300