How informative are explicit interests for recommender systems?

Nitya Shah

Research Master's Internship

Supervisor: Dr. Hannes Rosenbusch

Department of Psychological Methods

University of Amsterdam

December 2023

How informative are explicit interests for recommender systems?

How are platforms like TikTok and Instagram so engaging and gripping? How do its "For You Page" (FYP) pages churn out content that users are seemingly glued to? Zhang and Liu (2021) note how TikTok, in particular, goes much beyond being social media, but an entertainment platform where users consume content from random creators, on their popular and niche interests, all owing to their top-notch recommendation algorithm. TikTok's algorithm, as do several other social media recommender systems, uses various aspects of user behaviour and online social relations to tailor their recommendations (Zhang & Liu, 2021). For most of these apps, users do not need to explicitly specify or report their interests when they join or register. This leads to platforms heavily relying on their recommender systems, which are usually based on implicit measures such as how the user behaves on the app (for example, clicks, searches, repeat views, among others). Here, we test how explicit and implicit interests correlate on an entertainment platform similar to TikTok, but for opinion-sharing.

Recommender systems can use two kinds of user feedback: self-reported and behavioural. *Self-reported user feedback* is when users directly or explicitly inform the system what they thought of content that was recommended to them (for example, Instagram suggests content and sometimes asks users what they thought of it with "Interested" vs. "Not Interested" options) (Ricci et al., 2021). On the other hand, *behavioural user feedback* is when the users' actions on a platform are used to *infer* what they think of content recommended to them. This form of feedback is more implicit. Examples of this include liking or disliking a post, time spent watching a video, skipping videos, and others (Claypool et al., 2001; Núñez-Valdéz et al., 2012). Additionally, recommender systems and algorithms may sometimes suffer from the new user or *cold start problem*, where an initial lack of information about a user makes it difficult to suggest the first content they see (Freyne et al., 2009). In this case, collecting users' explicit interests (e.g., through a survey) could be a way to tackle this problem.

Explicit and Implicit Interests

Explicit interests are interests and preferences that are explicitly declared by the user and *implicit interests* are those which are *inferred* through the users' actions and behaviour. In the case of recommender systems, they infer an individual's interests through actions such as scrolling behaviour and dwelling or reading time, and use that to suggest content (Jayarathna & Shipman, 2017). At present, most recommender systems do not ask users what they are interested in and what content they want to see. They operate by suggesting content, and either ask users if they are interested in it or not (i.e., *self-reported user feedback*) or by inferring whether users enjoyed the content through their behaviour (i.e, *behavioural user feedback*).

An issue that features in a lot of previous literature (Jayarathna & Shipman, 2017) on recommender systems is that self-reported user feedback is not as easily accessible or available as behavioural feedback. One possible reason for this is that it requires more input and work on the users' part (Jannach et al., 2018). Furthermore, there is some ambiguity with self-reported user feedback: users could be rating the quality of the recommendation rather than how much they liked (or by extension, how interested they are) in the recommended content (Jannach et al., 2018). Consequently, in practice, behavioural feedback is more in use than self-reported user feedback for recommender systems. However, both forms of feedback, self-reported and behavioural, can be noisy (Jawaheer et al., 2010). Not all actions and behaviour indicate interest, they may just be accidental or natural movements. As for behavioural feedback, it tends to be more on the extreme positive or negative ends of the spectrum, as people rate something when they feel very strongly about it (Jawaheer et al., 2010). Thus, it could be the case that an initial (one-time) declaration of explicit interests is comparatively less arduous than self-reported user feedback that is asked for after several recommended items, making it less intrusive. There is a brief indication from previous literature which may suggest that users engage more with content aligning to their (explicit) interests. Di Gangi and Wasko (2016) note

that when users feel that interactions on social media are tailored to their interests, it leads to them engaging more. While this study looked at social interactions being tailored to personal interests, it may still suggest that users are likely to engage more when the content on social media is aligned with their interests.

In general, however, there is a gap in previous literature on the use of explicit interests for recommender systems, which is one of the things we address in our study. In this report, we specifically look at how explicit and implicit interests are correlated. Essentially, can users tell us what content they would like to see on entertainment platforms? Do they *really* enjoy content on topics that they claim they like?

Current Recommender Systems for Entertainment Platforms

We also looked at the features and components of recommender systems used by popular entertainment platforms to examine what the current state-of-the-art is and understand where the field is at. Our focus is on the algorithms of casual entertainment platforms (e.g., TikTok), considering that the app we conducted our study using is of a similar style, compared to platforms like Twitter and Facebook, where networking, as opposed to pure entertainment content, is more prevalent.

Narayanan (2023) briefly runs through some primary optimization targets that the different social media algorithms use: Twitter looks at users' interactions with tweets, YouTube uses watch time, and TikTok uses watch time, likes, comments and completion of a video. All these metrics are forms of behavioural feedback. Another consideration in recommendation algorithms is the exploration-exploitation trade off: recommending content that a user is known to like (exploitation) or also recommending different content to see if it appeals to a user (exploration). TikTok, for example, is known to be more on the exploration side of this spectrum. Jannach and Zanker (2022) suggest that discovering new content is a main part of the recommender system's job and leads to more users exhibiting greater engagement and interest. It could thus be the case that this contributes

to users engaging so much with TikTok, and its popularity.

TikTok's official website states that their recommender system uses user interactions (including likes, comments, shares, and content created), video information (such as captions, sounds, and hashtags) and device or account settings (which includes language and location) (TikTok, 2020). They place different weightage on different indicators, and then rank videos to gauge how much a user may like a certain piece of content before ultimately feeding it into a users' *For-You-Page* (TikTok, 2020). They also allow users to explicitly indicate that they are "Not Interested" in specific videos and hide or report certain content. Thus, TikTok also uses some form of self-reported user feedback. Narayanan (2023) and Wang (2022) note that there is a lack of information and research on the TikTok algorithm, however, which limits our knowledge on this matter. Specifically, not much is known about how TikTok sends novel and diverse content to their users' feeds. Since that is one of the unique features of TikTok, which is seemingly popular with users, it would be an important and interesting point to know.

TikTok's twin service in China, *Douyin*, initially uses basic user information, and data from a linked third-party social application if that has been used for login (Zhao, 2021). Further, as do several other social media platforms, Douyin uses *collaborative filtering*. Collaborative filtering, also known as "people-to-people correlation", suggests content to a user based on content liked by similar users (Ricci et al., 2021). This similarity between users is found through past behaviour and self-report feedback on a platform. Zhao (2021) suggests that recommendations made through collaborative filtering may allow users to discover new interests or those which they were not consciously aware of. As is the case with TikTok and Instagram reels (i.e., short videos), users can mark whether they like and dislike a video, and this updates Douyin's knowledge of the user (Zhao, 2021). In using collaborative feedback, Douyin too relies on behavioural feedback and has a form of self-reported user feedback as well, as in TikTok and Instagram. A similar trend of using both types of user feedback to personalise recommendations is seen

for YouTube too. Goodrow (2021), YouTube's VP of Engineering, notes that YouTube uses clicks, watch time, survey responses (rating the video), sharing, likes and dislikes to make recommendations.

As seen above, the recommender systems and algorithms of popular entertainment platforms *primarily* use different forms of behavioural feedback, as well as some self-reported user feedback. However, none of them use explicit interests, or ask users to mark their preferences at the time of joining the platform[1]. Given that explicit interests are not used in common social media recommender algorithms, as well as the gap in research in this field, it would be interesting to see whether users have insights into their own interests and if these insights are unbiased when they are probed explicitly. From a more practical perspective, collecting explicit interests at the very beginning may also benefit the cold start problem. Furthermore, Narayanan (2023)notes that the rate of engagement (i.e., the likelihood that a user engages with content recommended by the algorithm), with recommended content is merely 1% and in the exceptional case of TikTok, a bit over 5%. This also makes us question whether using explicit interests, and asking users upfront what they would like to see on the platform, influences how they interact with content we expect them to like.

Our Study

The goal of this project is to examine whether users' explicit interests correlate with their implicit interests on entertainment platforms. This was done by using data from a market research company which uses their own social media application to grow a user base. The app consists of a large user panel that answers opinion polls (in this report we will refer to these as *'polls'*), which are either created by a team at the company or other

[1] Reddit does ask for explicit interests at the time of joining, but from previous literature, they appear not to have a recommender system, which makes it difficult for us to consider Reddit when examining the current state of the field.

app users (for example, “If you have to dye your hair any colour from the rainbow, what colour would it be?”, Options: “Purple” / “Green” / “Orange” / “Other”). It essentially gamifies the act of opinion-sharing through coins and levels. Additionally, users can upvote or downvote a poll. When downvoting a poll, the user has to click on a reason for the downvote. It works a bit like TikTok in the sense that users can swipe up to go to the next piece of content: polls on the company’s social media app and videos on TikTok. In the company’s app, users can comment or upvote or downvote polls and similarly, on TikTok, click the heart to like a video or comment.

We aim to learn whether users’ explicit content preferences correlate with their implicit interests indicated by their in-app behaviour. More specifically, we examine if users’ positive implicit interest, indicated by upvoting polls matches their positive explicit interests (primary research question; analysis 1) and the converse, whether users’ negative implicit interests, indicated by downvoting polls, is correlated with their negative explicit interests (analysis 2). In addition to these, two other exploratory analyses were also performed. These examined whether users’ positive explicit interest and negative implicit interest correlate (analysis 3) and whether users’ negative explicit interest and positive explicit interest correlate (analysis 4). Analysis 2, 3 and 4 serve as alternative methods to assess our primary research question. With these, we aim to more precisely understand how strongly explicit and implicit interests correlate.

Hypotheses

Previous literature showed that users feel more engaged when social media interactions are tailored to their interests (Di Gangi & Wasko, 2016), which may indicate that they are more likely to upvote (i.e., exhibit a positive implicit interest) in content that aligns with their explicit interest. Furthermore, the well-established *theory of planned behaviour* (Ajzen, 1991) suggests that behavioural intentions are influenced by attitude towards the behaviour, societal norms and perceived behavioural control. In particular, it

suggests that having a more positive attitude towards the behaviour in question increases the intention to actually exhibit that behaviour. Applying this theory to our study would indicate that having a positive attitude towards a particular topic or engaging with it might make it more likely for a user to actually want to engage in a poll based on that topic (i.e., exhibit positive implicit behaviour). Similarly, the converse may hold as well: having a negative attitude towards a topic may result in less desire to engage with that topic. Given these indications from previous literature, we hypothesised the following H1 and its converse H2, corresponding to analysis 1 and 2.

H1. User's positive explicit interest (survey response indicating their interest in a topic) are positively correlated with their positive implicit interest in a topic (upvoting polls on that topic).

H2. User's negative explicit interest (survey response indicating whether they are disinterested in a topic) are positively correlated with their negative implicit interest in a topic (downvoting polls on that topic).

Method

Data Collection

Explicit Interests

We pulled out all responses to an interest survey that users get when they join the app. In the survey, users mark topics they would like to see more of (i.e. topics they are interested in) from a given list of topics and those they are not interested in. This resulted in a dataset consisting of 43848 users each rating 24 topics. However, not all of them upvoted or downvoted polls, and were thus excluded from the dataset. This interests survey consisted of two questions, translated here from German to English:

Q1. Which topics would you like to see more of?

Q2. Which topics do you not want to see?

Underneath both questions, users saw the following list of options: *Work, Vehicles, Games, Books/Reading, Food, Dating, Family, Film/Television, Friends/Relations, Money/Finance, Health, Internet/Technology, Cosmetics/Wellness, Fashion, Music, Politics, Travel, School, Spirituality, Sport, Animals/Nature, Celebrities/Gossip, Jokes/Humor, None*. Users were allowed to mark multiple topics (or None) for both questions. In our dataset, all user responses for interest in a topic and disinterest in a topic were dummy coded; for interests (Q1), 1 indicated a positive explicit interest in the topic and 0 indicated there was no explicit interest. Similarly, for disinterests (Q2), 1 indicated that the user had a negative explicit interest (i.e., was *explicitly disinterested*) in the topic, and 0 indicated that they were not explicitly disinterested in it. For each user, we thus had a record of topics they marked as their explicit interests and disinterests.

Implicit Interests

To measure implicit interests, we used the polls and related data from the company database. From the database, we first pulled out a set of all the polls that have been upvoted or downvoted *at least once*. This resulted in a set of 31270 polls that had been upvoted at least once and 7422 polls that had been downvoted at least once.

When a user comes across a poll on the app, they can answer it, skip it and swipe to the next, report it, upvote it if they like the poll or downvote it if not. In our analysis, we look at implicit interests in two ways - *positive implicit interest* through upvotes on polls and *negative implicit interest* through downvotes on polls. Initially, we planned to operationalise negative implicit interests by looking at the polls skipped by users. However, the company does not store this data given that it is a huge amount, which is why it was not possible to do the analysis using poll skips. Therefore, we decided to use only downvotes to operationalise negative implicit interest. More specifically, positive implicit interest for each user for a specific topic was defined as the proportion of topic-based polls upvoted out of the total number of upvotes given by this user. For example, for the topic "books", we look at the number of polls upvoted about "books" out of the total number of

polls upvoted by each user. The same is done for all topics, for each user. To measure negative implicit interest for each user, per topic, the same was done but with downvotes instead of upvotes. This was done because we performed all the analyses for *each* topic separately.

Analysis Plan

Due to certain issues that were discovered after looking at the company database and availability of data (primarily that poll skipping data was not there and thus could not be used), the analysis plan had to be altered a bit, and implicit interest was ultimately only operationalised using upvotes and downvotes.

Analysis 1. To examine the relationship between positive explicit interest and positive implicit interest for a specific topic (hypothesis 1), we performed a linear regression with explicit interest in the topic (1 for interest, 0 for not interested) as the independent variable and the proportion of upvotes for polls on that topic out of the total number of upvotes given by a user as the dependent variable.

Analysis 2. Next, to look at the relationship between negative explicit interest in a topic and negative implicit interest with that topic (hypothesis 2), we performed a linear regression with negative explicit interest in the topic (1 or 0)) as the independent variable and the proportion of downvotes for polls on that topic out of the total number of upvotes given by a user as the dependent variable.

Analysis 3. For our more exploratory analyses, to test the relationship between positive explicit interest in a topic and negative implicit interest, we again performed a linear regression with positive explicit interest in the topic (1 or 0) as the independent variable and the proportion of downvotes for polls on that topic out of the total number of downvotes given by a user as the dependent variable.

Analysis 4. Lastly, to investigate the relationship between negative explicit interest in a topic and positive implicit interest in that topic, we performed a linear regression with negative explicit interest in the topic (1 for disinterest, 0 for not

disinterested) as the independent variable and the proportion of upvotes for polls on that topic out of the total number of upvotes given by a user as the dependent variable.

Data Pre-processing

After pulling out the poll dataset and the user interests, several steps were undertaken to make the final dataset that would be used for the analysis. These steps included filtering the poll dataset and the user dataset in order to only retain users whose implicit and explicit interests we have. The exact details and pre-processing steps can be found in Appendix A.

Poll Labelling

The final poll dataset ($\approx$ 31,000 polls) consisted only of polls that at least one user had upvoted or downvoted. However, to perform the analyses, we also had to determine what topic each poll fell under. Assigning a topic label to 31,000 polls would have been extremely time consuming and inefficient if done manually. We thus decided to use OpenAI's GPT 3.5 to speed up this process. After several rounds of writing prompts and testing them on small sample datasets (the final prompt and code used is in Appendix B), we were also able to gauge what costs to expect if we were to use GPT 3.5 for 31,000 polls. Due to cost constraints on the company's end, and rate limits and server timeouts on OpenAI's end, it was not possible to label all polls using GPT. Ultimately, it was only possible to label 2000 polls (1000 upvoted and 1000 downvoted polls) using GPT 3.5. Furthermore, since all the polls were in German, it was not possible to manually give topic labels to polls, as non-German speakers. We compared the topic labels assigned to a sample set of polls by an external native German speaker and a non-German speaker (author 1, NS), which revealed several discrepancies. This was attributed to some ambiguous translations from German to English using both Google Translate and DeepL, and a lack of colloquial context.

The prompt for this task asked GPT to label a poll as 'Other' if it did not know

which other category the poll best fits in. GPT marked 256 upvoted polls and 294 downvoted polls as 'Other'. These polls were then manually re-labelled by a native German speaker, after which 172 of the 256 upvoted polls and 180 of the 294 downvoted polls were assigned a topic label. The rest of the polls marked as 'Other' by GPT and by a native German speaker were thus filtered out and excluded from the dataset.

A final visual inspection was done to check that all polls were assigned to a topic from the pre-defined topic list. If this was not the case, and the poll could not be manually re-labelled it was removed. After all these manual checks were completed, the dataset consisted of 913 upvoted polls (i.e., polls that were upvoted at least once) and 885 downvoted polls (i.e., polls that were downvoted at least once).

Final Dataset

Some additional steps were taken to prepare the final dataset. The details can be found in Appendix C.

Sample Characteristics

The final dataset for the positive implicit interest analyses consisted of 385 users and for the negative implicit interest analyses, 270 users. Table 1 further shows the number of users in each analysis for each topic.

Results

The analyses were then performed for each topic separately. This resulted in 24 separate sub-analyses for each of the four planned analyses. Due to this, the sample size of each sub-analysis varied (e.g., there were 214 users who upvoted at least one poll about food, while only 25 did so for polls on celebrities & gossip). Table 1 consists of the number of users that exhibited a positive implicit interest for each topic (by upvoting polls on that topic) or a negative implicit interest (by downvoting polls on that topic). For analysis 1 and 4, refer to the sample size for the 'positive implicit interest' column in Table 1 and for

analysis 2 and 3, refer to the sample size in the 'negative implicit interest' column.

Before undertaking the analyses, assumption checks were performed for all regression models. The assumptions of normality and homoscedasticity were checked through visual inspection; while the normality assumption was not violated in most cases, it was not as straightforward for the homoscedasticity assumption. However, since we had multiple replications for each analysis (i.e., one model for each topic, for each analysis, thus leading to 24 replications per analysis), these assumptions are not problematic for us.

Table 1

Number of users that exhibited a positive and negative implicit interest for each topic (by upvoting or downvoting polls, respectively).

	Users (N) Positive implicit interest	Users (N) Negative implicit interest
Food	214	69
Animals & Nature	168	42
Vehicles	69	13
Games	70	11
Books & Reading	17	10
Dating	69	48
Family	89	23
Movies & Television	113	33
Friends & Relations	141	25
Money & Finance	65	19
Health	155	53
Internet & Technology	145	42
Cosmetics & Wellness	99	19
Fashion	113	8
Music	55	9
Politics	110	41
Celebrities & Gossip	25	2
Travel	93	22
School	60	21
Shopping	86	8
Spirituality	83	21
Sport	88	28
Jokes & Humor	57	40
Work	70	10

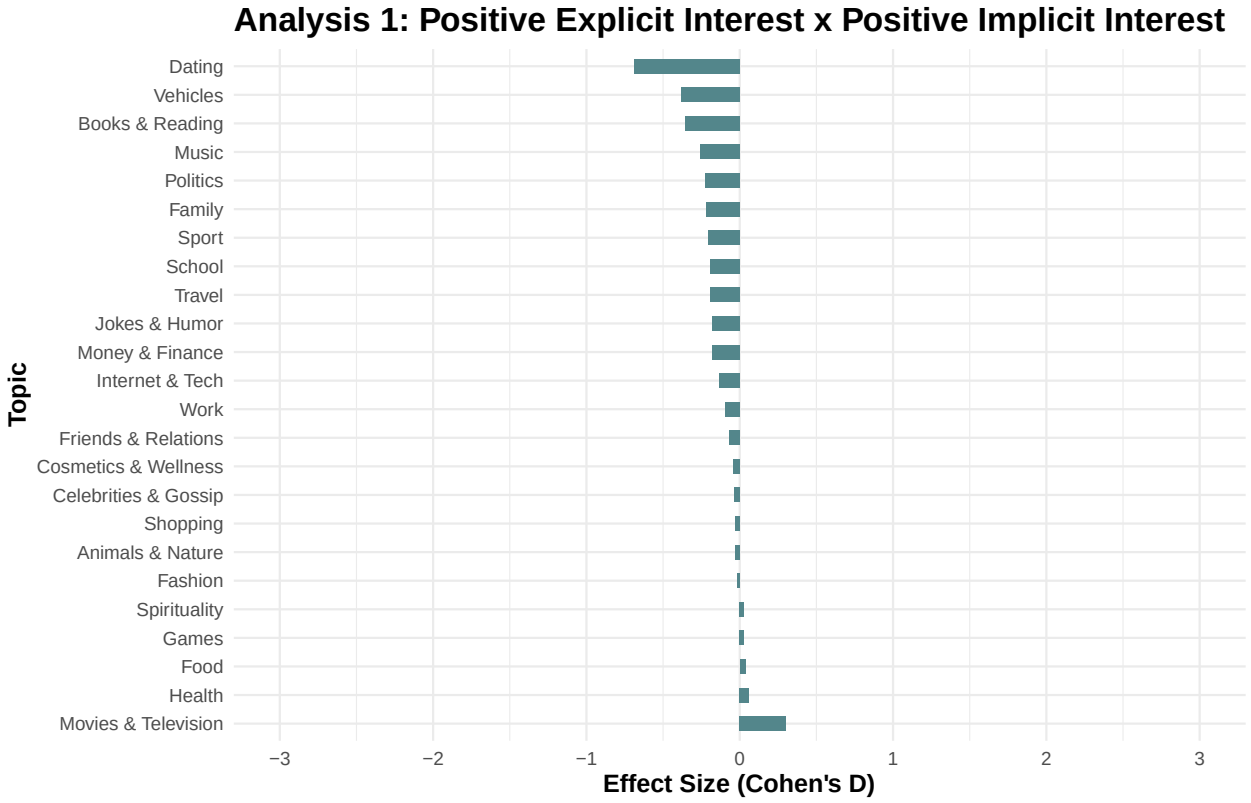
Analysis 1: Positive explicit interest and positive implicit interest

Analysis 1 (results in Table 2) examined the relationship between positive explicit interest in a topic and positive implicit interest, in the form of upvoting a poll. More specifically, we wanted to check if there was any significant difference in the positive implicit interest between those who are explicitly interested in a topic and those who are not. A significant difference was only found for the topic dating. Other than that, there were no notable differences between groups for any of the 24 topics, which indicates that positive explicit interest in a topic may not be correlated with positive implicit interest.

Figure 1 summarises the results of analysis 1 for all 24 topics, by comparing the effect sizes across topics. A negative effect size (Cohen's D), in this case, shows that those with a positive explicit interest in the topic exhibit higher positive implicit interest, measured by the proportion of upvotes.

Figure 1

Effect sizes for analysis 1



Note. *A* negative effect size (Cohen's D) indicates that the mean positive implicit interest (proportion of upvotes for the topic) for the group which had a positive explicit interest (group 1) is greater than those without an explicit interest (group 0) (i.e., group 1 mean > group 0 mean). A positive effect size indicates the converse.

Table 2

Analysis 1: Positive explicit interest and positive implicit interest

	β	t	F	R^2	p
Food	-0.01	-0.25	$F(1, 212) = 0.06$	0.000	0.802
Animals & Nature	0.01	0.18	$F(1, 166) = 0.03$	-0.01	0.855
Vehicles	0.07	1.45	$F(1, 67) = 2.10$	0.02	0.152
Games	-0.01	-0.12	$F(1, 68) = 0.01$	-0.01	0.905
Books & Reading	0.01	0.70	$F(1, 15) = 0.49$	-0.03	0.497
Dating	0.14	2.46	$F(1, 67) = 6.06$	0.07	0.016*
Family	0.04	1.02	$F(1, 87) = 1.03$	0.00	0.313
Movies & Television	-0.07	-1.54	$F(1, 111) = 2.36$	0.01	0.127
Friends & Relations	0.02	0.40	$F(1, 139) = 0.16$	-0.01	0.687
Money & Finance	0.06	0.72	$F(1, 63) = 0.52$	-0.01	0.474
Health	-0.014	-0.35	$F(1, 153) = 0.12$	-0.006	0.730
Internet & Technology	0.03	0.78	$F(1, 143) = 0.60$	-0.003	0.439
Cosmetics & Wellness	0.01	0.21	$F(1, 97) = 0.04$	-0.01	0.834
Fashion	0.00	0.09	$F(1, 111) = 0.01$	-0.01	0.930
Music	0.07	0.93	$F(1, 53) = 0.87$	-0.00	0.357
Politics	0.05	1.17	$F(1, 108) = 1.38$	0.00	0.244
Celebrities & Gossip	0.004	0.08	$F(1, 23) = 0.01$	-0.04	0.940
Travel	0.04	0.88	$F(1, 91) = 0.77$	-0.00	0.38
School	0.04	0.48	$F(1, 58) = 0.23$	-0.01	0.633
Shopping	0.01	0.15	$F(1, 84) = 0.02$	-0.01	0.882
Spirituality	-0.01	-0.11	$F(1, 81) = 0.01$	-0.01	0.915
Sport	0.05	0.94	$F(1, 86) = 0.89$	-0.00	0.348
Jokes & Humor	0.04	0.68	$F(1, 55) = 0.47$	-0.01	0.498
Work	0.01	0.35	$F(1, 68) = 0.12$	-0.01	0.725

*$p < .05$.

Analysis 2: Negative Explicit Interest and Negative Implicit Interest

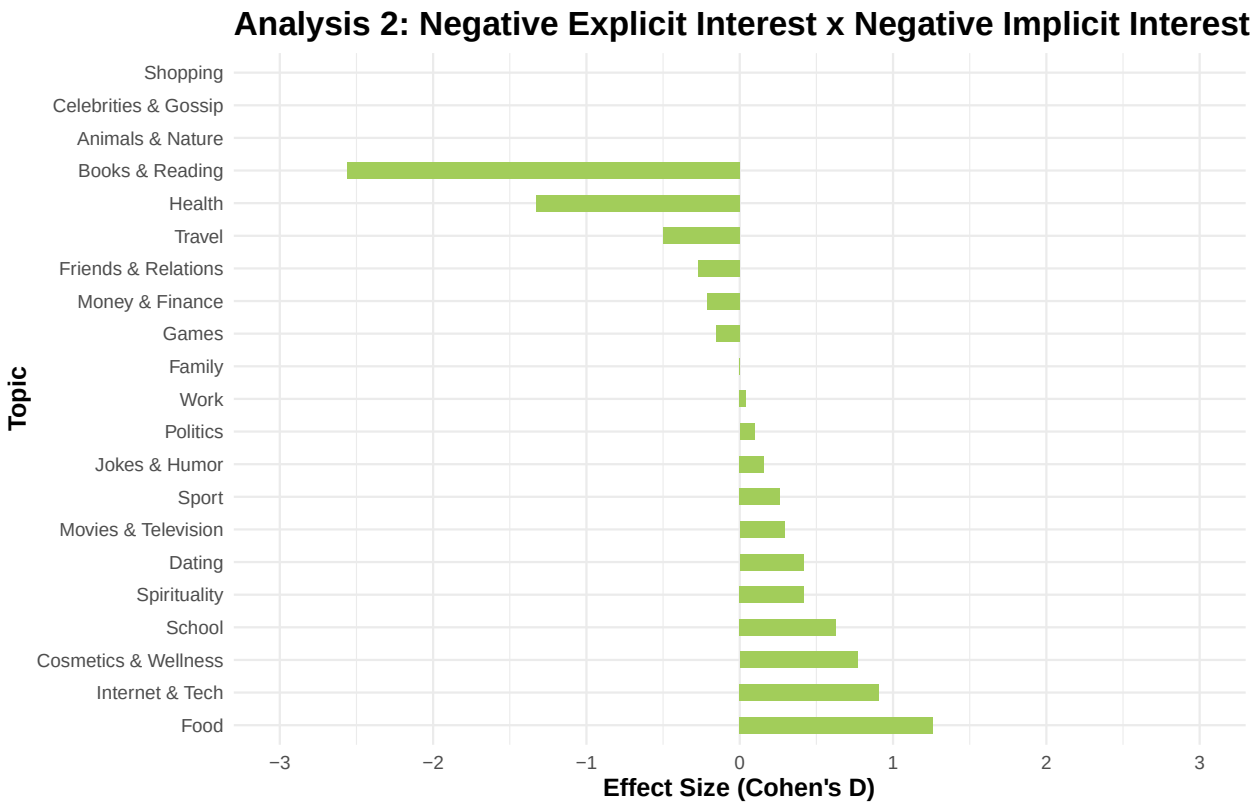
Next, analysis 2 (results in Table 3) looked at the relationship between negative explicit interest and negative implicit interest. Significant differences between those with an explicit disinterest in the topic and those who are not were found for topics like food, books & reading and health. However, it should be noted that there is just one user who is explicitly disinterested in health and in books & reading, which makes it extremely hard to draw a concrete conclusion. In sum, again, negative explicit interest in a topic does not seem to be correlated with negative implicit interest.

For the topics music, vehicles, and fashion, of the users who downvoted poll(s) on these topics, none marked that they are explicitly disinterested in these topics, and this might have resulted in an NA since there is no comparison group. However, this in itself is an interesting result which goes against our hypothesis as those who are (positively) explicitly interested in the topic exhibit a negative implicit interest for topics that they would perhaps be expected to have a positive implicit interest in.

Figure 2 summarises the results of analysis 2 for all 24 topics, by comparing the effect sizes across topics. A negative effect size (Cohen's D), in this case, shows that those with a negative explicit interest (or explicit disinterest, essentially) in the topic exhibit greater negative implicit interest, measured by the proportion of downvotes.

Figure 2

Effect sizes for analysis 2



Note. A negative effect size (Cohen's D) indicates that the mean negative implicit interest (proportion of downvotes for the topic) for the group which had a negative explicit interest (i.e., explicit disinterest) (group 1) is greater than those without an explicit disinterest (group 0) (i.e., group 1 mean > group 0 mean). A positive effect size indicates the converse.

The effect sizes for shopping, celebrities & gossip and animals & nature produced an NA. Furthermore, the analyses themselves for music, vehicles and fashion produced an NA, which is why they are not present in the plot at all.

Table 3

Analysis 2: Negative Explicit Interest and Negative Implicit Interest

	β	t	F	R^2	p
Food	-0.43	-2.13	$F(1, 67) = 4.55$	0.05	0.037*
Animals & Nature	0.60	1.99	$F(1, 40) = 3.98$	0.07	0.053
Vehicles	NA	NA	NA	NA	NA
Games	0.05	0.23	$F(1, 9) = 0.05$	-0.10	0.826
Books & Reading	0.78	3.24	$F(1, 8) = 10.47$	0.51	0.012*
Dating	-0.14	-1.42	$F(1, 46) = 2.02$	0.02	0.162
Family	0.001	0.01	$F(1, 21) = 4.069e - 05$	-0.05	0.995
Movies & Television	-0.12	-0.65	$F(1, 31) = 0.42$	-0.02	0.523
Friends & Relations	0.08	0.44	$F(1, 23) = 0.19$	-0.03	0.667
Money & Finance	0.07	0.37	$F(1, 17) = 0.14$	-0.05	0.714
Health	0.49	2.24	$F(1, 51) = 4.999$	0.07	0.030*
Internet & Technology	-0.35	-1.73	$F(1, 40) = 2.98$	0.05	0.092
Cosmetics & Wellness	-0.11	-1.55	$F(1, 17) = 2.41$	0.07	0.139
Fashion	NA	NA	NA	NA	NA
Music	NA	NA	NA	NA	NA
Politics	-0.04	-0.26	$F(1, 39) = 0.07$	-0.02	0.797
Celebrities & Gossip	0.49	NaN	NaN	NaN	NaN
Travel	0.17	0.80	$F(1, 20) = 0.64$	-0.02	0.432
School	-0.22	-1.36	$F(1, 19) = 1.85$	0.04	0.190
Shopping	-0.23	-0.61	$F(1, 6) = 0.38$	-0.10	0.562
Spirituality	-0.18	-0.93	$F(1, 19) = 0.87$	-0.01	0.364
Sport	-0.10	-0.60	$F(1, 26) = 0.36$	-0.02	0.552
Jokes & Humor	-0.06	-0.33	$F(1, 38) = 0.11$	-0.02	0.745
Work	-0.02	-0.06	$F(1, 8) = 0.00$	-0.12	0.950

Note. A NaN was produced since there were too few data points (i.e., 2.)

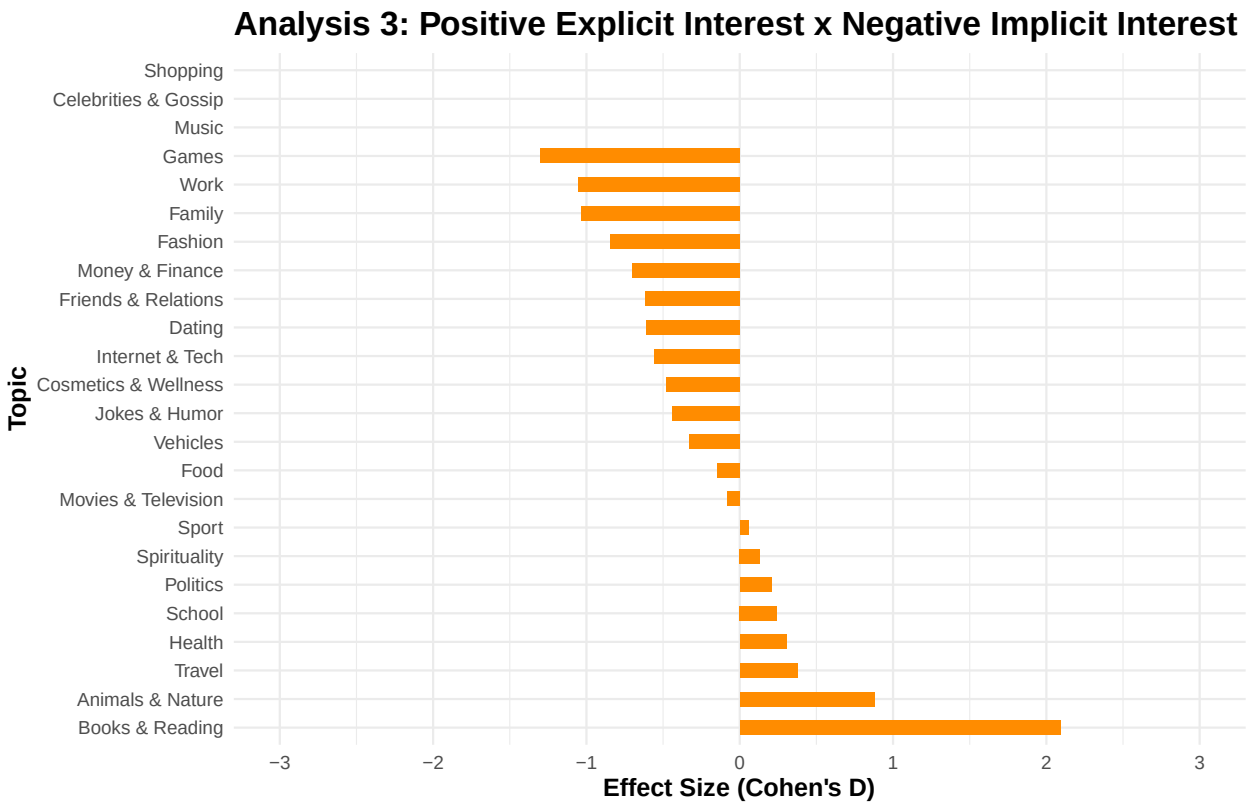
Analysis 3: Positive Explicit Interest and Negative Implicit Interest

Analysis 3 (results in Table 4) specifically examined the relationship between positive explicit interest and negative implicit interest. Significant differences between those who are explicitly interested and those who are not are only seen for the following topics: animals & nature, family and music. Thus, there was a difference in the number of downvotes given by users who said they are explicitly interested in animals & nature, family or music compared to users who said they are not. However, there was no consistent direction of the effect. For example, for the "family" topic, users who said they were explicitly interested in the topic gave more downvotes, but the converse was seen for "music"; users who said they were not explicitly interested in the topic gave more downvotes. This makes it difficult to draw a robust conclusion, despite the significant effects. It should also be noted that for each of these topics, the analysis is characterised by a higher degree of uncertainty due to the small sample sizes of downvoted polls. Considering that significant differences were only found for very few topics, with no robust conclusion and issues with sample size, it can be concluded that positive explicit interest may not be very indicative of negative implicit interest either.

Figure 3 summarises the results of analysis 3 for all 24 topics, by again comparing the effect sizes across topics. A negative effect size (Cohen's D), here, shows that those with a positive explicit interest in the topic exhibit higher negative implicit interest, measured by the proportion of downvotes.

Figure 3

Effect sizes for analysis 3



Note. *A* negative effect size (Cohen's D) indicates that the mean negative implicit interest (proportion of downvotes for the topic) for the group which had a positive explicit interest (group 1) is greater than those without an explicit interest (group 0) (i.e., group 1 mean > group 0 mean). A positive effect size indicates the converse.

The effect sizes could not be calculcated for shopping, celebrities & gossip and music, they simply produced NAs.

Table 4

Analysis 3: Positive Explicit Interest and Negative Implicit Interest

	β	t	F	R^2	p
Food	0.05	0.60	$F(1, 67) = 0.36$	0.01	0.552
Animals & Nature	-0.26	-2.34	$F(1, 40) = 5.47$	0.10	0.025*
Vehicles	0.12	0.58	$F(1, 11) = 0.34$	-0.06	0.574
Games	0.35	1.66	$F(1, 9) = 2.76$	0.15	0.131
Books & Reading	-0.64	-3.24	$F(1, 8) = 10.51$	0.51	0.012
Dating	0.21	1.84	$F(1, 46) = 3.37$	0.05	0.073
Family	0.34	2.18	$F(1, 21) = 4.74$	0.15	0.041*
Movies & Television	0.03	0.23	$F(1, 31) = 0.05$	-0.03	0.817
Friends & Relations	0.18	1.53	$F(1, 23) = 2.33$	0.05	0.141
Money & Finance	0.23	1.51	$F(1, 17) = 2.27$	0.07	0.151
Health	-0.12	-1.11	$F(1, 51) = 1.22$	0.004	0.274
Internet & Technology	0.22	1.81	$F(1, 40) = 3.27$	0.05	0.078
Cosmetics & Wellness	0.07	1.04	$F(1, 17) = 1.08$	0.00	0.314
Fashion	0.28	1.16	$F(1, 6) = 1.34$	0.05	0.291
Music	-0.80	-4.59	$F(1, 7) = 21.08$	0.72	0.003**
Politics	-0.08	-0.61	$F(1, 39) = 0.38$	-0.02	0.544
Celebrities & Gossip	-0.49	NaN	NaN	NaN	NaN
Travel	-0.13	-0.87	$F(1, 20) = 0.75$	-0.01	0.396
School	-0.09	-0.44	$F(1, 19) = 0.19$	-0.04	0.669
Shopping	-0.39	-1.09	$F(1, 6) = 1.19$	0.03	0.317
Spirituality	-0.06	-0.26	$F(1, 19) = 0.07$	-0.05	0.799
Sport	-0.02	-0.13	$F(1, 26) = 0.02$	-0.04	0.900
Jokes & Humor	0.16	1.39	$F(1, 38) = 1.92$	0.02	0.174
Work	0.43	1.53	$F(1, 8) = 2.35$	0.13	0.164

*$p < .05$. **$p < .01$.

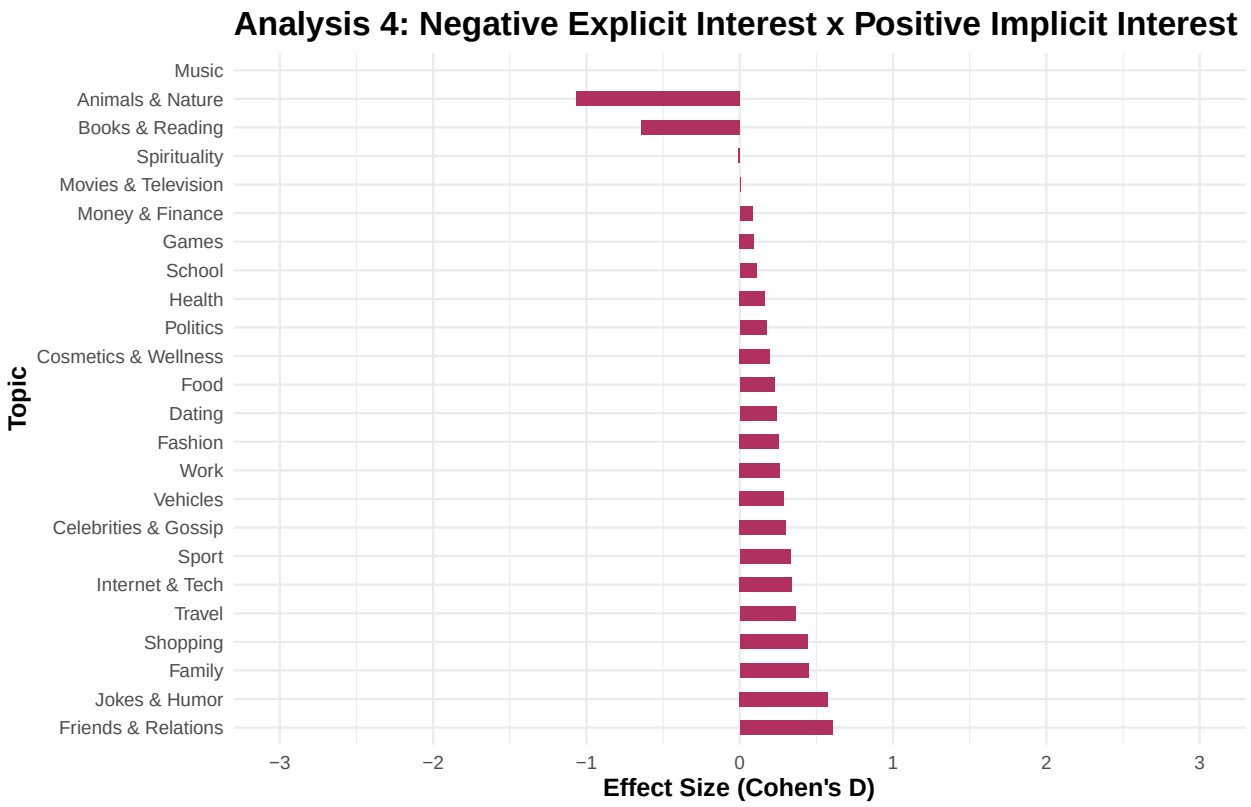
Analysis 4: Negative Explicit Interest and Positive Implicit Interest

Lastly, analysis 4 (results in Table 5) looked at the relationship between a negative explicit interest (i.e., explicit disinterest) in a topic and positive implicit interest, in the form of upvoting. More specifically, we wanted to check if there was any significant difference in upvoting behaviour between those explicitly disinterested in a specific topic and those who are not. No significant differences were found between these two groups, which indicates that explicit disinterest in a topic may not be correlated with a positive implicit interest in a topic either.

Figure 4 summarises the results of analysis 4 for all 24 topics, by comparing the effect sizes across topics. A negative effect size (Cohen's D), in this case, shows that those with a negative explicit interest (or explicit disinterest, essentially) in the topic exhibit greater positive implicit interest, measured by the proportion of upvotes.

Figure 4

Effect sizes for analysis 4



Note. A negative effect size (Cohen's D) indicates that the mean positive implicit interest (proportion of upvotes for the topic) for the group which had a negative explicit interest (i.e., explicit disinterest) (group 1) is greater than those without an explicit disinterest (group 0) (i.e., group 1 mean > group 0 mean). A positive effect size indicates the converse.

The effect size could not be calculated for music.

Table 5

Analysis 4: Negative Explicit Interest and Positive Implicit Interest

	β	t	F	R^2	p
Food	-0.05	-0.70	$F(1, 212) = 0.49$	-0.00	0.486
Animals & Nature	0.30	1.50	$F(1, 166) = 2.25$	0.01	0.136
Vehicles	-0.05	-1.09	$F(1, 67) = 1.20$	0.00	0.278
Games	-0.03	-0.29	$F(1, 68) = 0.08$	-0.01	0.774
Books & Reading	0.02	0.86	$F(1, 15) = 0.73$	-0.02	0.406
Dating	-0.05	-0.997	$F(1, 67) = 0.99$	-9.556e-05	0.322
Family	-0.08	-0.76	$F(1, 87) = 0.58$	-0.00	0.447
Movies & Television	-0.001	-0.02	$F(1, 111) = 0.00$	-0.01	0.988
Friends & Relations	-0.14	-1.66	$F(1, 139) = 2.76$	0.01	0.099
Money & Finance	-0.03	-0.20	$F(1, 63) = 0.04$	-0.02	0.846
Health	-0.04	-0.23	$F(1, 153) = 0.05$	-0.006	0.816
Internet & Technology	-0.08	-1.58	$F(1, 143) = 2.50$	0.01	0.116
Cosmetics & Wellness	-0.04	-0.80	$F(1, 97) = 0.63$	-0.00	0.429
Fashion	-0.04	-1.09	$F(1, 111) = 1.18$	0.00	0.280
Music	-0.11	-0.39	$F(1, 53) = 0.152$	-0.02	0.699
Politics	-0.04	-0.75	$F(1, 108) = 0.56$	-0.00	0.455
Celebrities & Gossip	-0.04	-0.72	$F(1, 23) = 0.52$	-0.02	0.478
Travel	-0.08	-1.36	$F(1, 91) = 1.85$	0.01	0.178
School	-0.02	-0.42	$F(1, 58) = 0.17$	-0.01	0.679
Shopping	-0.08	-1.26	$F(1, 84) = 1.58$	0.01	0.212
Spirituality	0.00	0.05	$F(1, 81) = 0.00$	-0.01	0.958
Sport	-0.08	-1.06	$F(1, 86) = 1.13$	0.00	0.291
Jokes & Humor	-0.12	-0.97	$F(1, 55) = 0.94$	-0.00	0.337
Work	-0.04	-0.90	$F(1, 68) = 0.80$	-0.00	0.373

The results do not support any of our hypotheses as there seems to be almost no indication that users' positive explicit interest is correlated with positive implicit interest or the converse, that negative explicit interest is correlated with negative implicit interest. Analyses 2 and 3, which were more exploratory, also did not indicate that there is a significant difference in upvoting and downvoting behaviour between users who report that they have an explicit interest or disinterest in a topic.

Discussion

The current state-of-the-art recommender systems on social media and entertainment platforms have all primarily been using behavioural user feedback (for example, viewing time, scrolling behaviour, video completion). Some add scarce self-reported user feedback (for example, asking users if they are "Interested" or "Not Interested" in a suggested post), but generally such explicit measures are less popular due to them requiring more effort and input on the part of the user (Jannach et al., 2018).

Contrary to our expected results, we found that a positive explicit interest in a certain topic (e.g., sports, fashion, books) is not associated with positive implicit interest (more upvoting behaviour), and that a negative explicit disinterest in a certain topic is not associated with negative implicit interest (more downvoting behaviour). We did not find the converse either: being explicitly interested in a topic was not correlated with negative implicit interest (more downvoting), and being explicitly disinterested in a topic was not correlated with positive implicit interest (more upvoting). Our primary and the conceptual replications did not support our expectations.

Our findings, indicating no relationship between explicit and implicit interests, do not overall align with findings from literature on the broader psychological constructs of explicit and implicit attitudes. Hofmann et al. (2005) and Phipps et al. (2019) conducted meta-analyses where they found a significant correlation between explicit and implicit attitudes, although small. The MODE (Motivation and Opportunity as Determinants of

the attitude-behaviour relation) model (Fazio, 1990) adds some nuance to this finding, suggesting that explicit (self-reported) attitudes may be influenced by motives such as social desirability (Phipps et al., 2019). For example, one's age and understanding of societal issues and prejudices influences their explicit attitude (Raabe & Beelmann, 2011), emphasising that the relationship's nature and strength depend on other motives (Phipps et al., 2019). These findings indicate a complex relationship between explicit and implicit attitudes, and that while they sometimes align, there are cases where they differ. As mentioned above, a key factor that influences self-reports or explicit attitudes is the social desirability bias. This leads to participants answering survey questions in a socially desirable manner and conform to norms, in order to maintain a societally accepted image of themselves (Krumpal, 2013). The social desirability bias may play a more significant role when reporting attitudes on sensitive topics, such as societal issues (Gawronski & De Houwer, 2014). In our study, asking users something less controversial such as their interests may have had a comparatively lower influence of social desirability bias on self-reported on explicit interests. Therefore, we may expect to see a greater discord and difference between explicit and implicit attitudes regarding sensitive societal topics, those which may be perceived as intrusive, risky to answer, or more prone to the social desirability bias (Krumpal, 2013; Olsson, 2023). This could include topics such as race, sex/gender, crime and politics, among others. These also often have critical consequences such as racial, ethnic or gender biases in the hiring process or during elections and voting. For example, on one hand, Hooghe and Reeskens (2007) found that populist radical-right (PRR) parties gained more votes and support in the actual elections than was projected in voter surveys, while John and Margetts (2009) found that implicit support for PRRs was greater than explicit support through votes (Bos et al., 2018). These findings too reiterate that sometimes explicit and implicit attitudes do diverge, especially when there is a greater motivation or reason to control either one's explicit support or implicit support, depending on the context.

Although our findings do not corroborate with previous psychological literature on explicit and implicit attitudes, there are some important caveats to discuss and take note of. A key difference between our study and most studies in this field are how implicit attitudes are operationalised. We look at behaviour (i.e., a user upvoting and downvoting polls), to measure implicit interests (or attitudes), but several studies in psychological research use some form of test (e.g., the Implicit Association Test (IAT)). Some of these studies examine implicit attitudes, explicit attitudes and explicit behaviour as different constructs. They also specifically look at whether implicit attitudes can be predictive of behaviour. However, we consider behaviour in itself an indicator of implicit interest in our study. This makes it a bit more complex to compare our results with previous literature, as it questions whether we are indeed measuring the same construct as that measured in past research. Furthermore, there has been criticism surrounding the validity, reliability and predictive power of implicit measures of attitude (Dai & Albarracín, 2022). Our design, on the contrary, examines implicit interest or attitudes in a natural setting by looking at behaviour, thus lending to greater ecological validity. In sum, psychological research could adopt more ecologically valid measures of implicit attitudes, but it is also not straightforward to compare our results and previous literature considering how we operationalise our constructs.

As for practitioners and businesses, it is not very clear if using explicit interests to mitigate the cold start problem works yet. Our study indicates that explicit interests are not informative for recommender systems, but adjustments can be made to our research design in order to be more sure of this. It could also be interesting for businesses and platforms to test the effectiveness and examine the user experience of including self-reported user feedback (i.e., users being asked if they are "Interested" or "Not Interested" in a new piece of content recommended to them). Even if an explicit measure such as explicit interests may not be effect, it could be the case that other measures like self-reported user feedback has indeed been helping the system better learn the users'

interests.

However, for researchers, it is still very interesting to understand whether users can report valid interests and if there is a difference between users. A future research line could also investigate whether there is a difference between users in reporting valid interests, and factors that influence the alignment of some users' explicit and implicit interests.

Limitations and directions for future research

As mentioned earlier, we cannot conclude with certainty that explicit interests are not informative and useful for predicting user engagement. Changing certain aspects of our design may allow us to make firmer conclusions. Operationalising implicit interests using upvotes and downvotes may miss more subtle indicators (for example, time spent on a poll, bookmarking or sharing a poll, and poll skipping). Upvoting and downvoting (and giving a reason for the downvote) of polls are also *explicit* indicators of interest in some way, and thus may be susceptible to the sparsity issue that self-reported user feedback faces. Additionally, it is possible that only those who feel very strongly, in either the positive or negative direction, may give an upvote or downvote (Jawaheer et al., 2010). Essentially, users may be explicitly or implicitly interested in certain content but may not express their like or dislike for it through upvotes and downvotes. Lastly, the social desirability bias may also influence our operationalisation of implicit interests. App users could be conscious of the polls they upvote or downvote knowing that this data is accessible to the company, and could perhaps be refraining from acting as they would in an environment free of surveillance. These issues suggest that our design may benefit from alternative operationalisations of engagement, especially for the implicit interest indicators.

While rethinking our design might help more clearly understand how informativer explicit interests are for recommender systems, there are some concerns and considerations with regards to using explicit interests in practice. The most basic question is how likely would people be to declare their explicit interests and content preferences when they join

an app? Previous literature only discusses issues pertaining to collecting self-reported user feedback, but not explicit interests, indicating that more future research could also investigate it or use it, as we did in this study. While collecting self-reported user feedback is more arduous and may break the user's flow and change their usual functioning on the app (Palanivel & Sivakumar, 2010), a one-time declaration of explicit interests may be less intrusive and time-consuming. Other issues with regards to users marking their explicit interests when they join or register for an app are that the user may answer quickly just to be able to use the platform, and their interests may change or evolve over time (Alrehili et al., 2022).

Conclusion

To conclude, our study did not yield any significant conclusions with regards to how explicit and implicit interests correlate. Thus, for now, it suggests that explicit interests are not very informative for recommender systems. However, with some changes in the research design, we may get a clearer picture whether current recommender systems could benefit from explicitly asking users what content they like. For psychological research, our findings indicate that there is some dissonance between explicit and implicit attitudes, which has implications for key issues such as societal biases, politics, and law, among others.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Alrehili, M. M., Yafooz, W. M., Alsaeedi, A., Emara, A.-H. M., Saad, A., & Al Aqrabi, H. (2022). The impact of personality and demographic variables in collaborative filtering of user interest on social media. *Applied Sciences*, *12*(4), 2157. https://doi.org/10.3390/app12042157

Bos, L., Sheets, P., & Boomgaarden, H. G. (2018). The role of implicit attitudes in populist radical-right support. *Political Psychology*, *39*(1), 69–87. https://doi.org/10.1111/pops.12401

Claypool, M., Le, P., Wased, M., & Brown, D. (2001). Implicit interest indicators. *Proceedings of the 6th International Conference on Intelligent Iser Interfaces*, 33–40.

Dai, W., & Albarracín, D. (2022). It's time to do more research on the attitude–behavior relation: A commentary on implicit attitude measures. *Wiley Interdisciplinary Reviews: Cognitive Science*, *13*(4), e1602. https://doi.org/10.1002/wcs.1602

Di Gangi, P. M., & Wasko, M. M. (2016). Social media engagement theory: Exploring the influence of user engagement on social media usage. *Journal of Organizational and End User Computing (JOEUC)*, *28*(2), 53–73. https://doi.org/10.4018/joeuc.2016040104

Freyne, J., Jacovi, M., Guy, I., & Geyer, W. (2009). Increasing engagement through early recommender intervention. *Proceedings of the Third ACM Conference on Recommender Systems*, 85–92.

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 283–310). Cambridge University Press. 10.1017/CBO9780511996481.016

Goodrow, C. (2021). On youtube's recommendation system. blog youtube. https://blog.youtube/inside-youtube/on-youtubes-recommendation-system/

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality Social Psychology Bulletin*, *31*(10), 1369–1385. https://doi.org/10.1177/0146167205275613

Jannach, D., Lerche, L., & Zanker, M. (2018). Recommending based on implicit feedback. In *Social information access: Systems and technologies* (pp. 510–569). Springer. https://doi.org/10.1007/978-3-319-90092-6

Jannach, D., & Zanker, M. (2022). Value and impact of recommender systems. In *Recommender systems handbook* (pp. 519–546). Springer. https://doi.org/10.1007/978-1-0716-2197-4_14

Jawaheer, G., Szomszor, M., & Kostkova, P. (2010). Comparison of implicit and explicit feedback from an online music recommendation service. *Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems*, 47–51. https://doi.org/10.1145/1869446.1869453

Jayarathna, S., & Shipman, F. (2017). Analysis and modeling of unified user interest. *2017 IEEE International Conference on Information Reuse and Integration (IRI)*, 298–307. https://doi.org/10.1109/iri.2017.46

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, *47*(4), 2025–2047. https://doi.org/10.1007/s11135-011-9640-9

Narayanan, A. (2023). Understanding social media recommendation algorithms. https://knightcolumbia.org/content/understandingsocial-media-recommendation-algorithms

Núñez-Valdéz, E. R., Lovelle, J. M. C., Martínez, O. S., García-Díaz, V., De Pablos, P. O., & Marín, C. E. M. (2012). Implicit feedback techniques on recommender systems

applied to electronic books. *Computers in Human Behavior*, *28*(4), 1186–1193. https://doi.org/10.1016/j.chb.2012.02.001

Olsson, F. (2023). The effect of implicit racial bias on right-wing populist support. *French Politics*, *21*(1), 81–103. https://doi.org/10.1057/s41253-022-00201-0

Palanivel, K., & Sivakumar, R. (2010). A study on implicit feedback in multicriteria e-commerce recommender system. *Journal of Electronic Commerce Research*, *11*(2).

Phipps, D. J., Hagger, M. S., & Hamilton, K. (2019). A meta-analysis of implicit and explicit attitudes in children and adolescents.

Ricci, F., Rokach, L., & Shapira, B. (2021). Recommender systems: Techniques, applications, and challenges. *Recommender Systems Handbook*, 1–35. https://doi.org/10.1007/978-1-0716-2197-4_1

TikTok. (2020). How tiktok recommends videos #foryou. *Retrieved December*, *12*, 2022. https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you

Wang, P. (2022). Recommendation algorithm in tiktok: Strengths, dilemmas, and possible directions. *International Journal of Social Science Studies*, *10*, 60. https://doi.org/10.11114/ijsss.v10i5.5664

Zhang, M., & Liu, Y. (2021). A commentary of tiktok recommendation algorithms in mit technology review 2021. *Fundamental Research*, *1*(6), 846–847. https://doi.org/10.1016/j.fmre.2021.11.015

Zhao, Z. (2021). Analysis on the "douyin (tiktok) mania" phenomenon based on recommendation algorithms. *E3S Web of Conferences*, *235*, 03029. https://doi.org/10.1051/e3sconf/202123503029

Appendix A

Data pre-processing steps

After pulling out the poll dataset and the user interests, several steps were undertaken to prepare the final dataset that would be used for the analysis.

As mentioned earlier, the user interests dataset consisted of topics that each user marked as their explicit interests or disinterests, using binary coding (0s or 1s). Further detail can be found in the Method section, under Explicit Interests.

Next, the poll dataset consisted of the IDs of users who had upvoted or downvoted each poll. We first extracted the list of all these user IDs and filtered the unique user IDs. Using that list of unique user IDs, we first filtered the interests survey to retain only the users who had upvoted or downvoted at least one poll. This was done so that we have the implicit interests of each user that are part of our final dataset and analysis.

Lastly, we filtered the poll dataset such that it only consisted of polls that were upvoted or downvoted by those in our final user set (i.e., users who filled the interests survey. Thus, those users whose explicit interests we have). After these steps. we ended up with a dataset of 2833 users, and approximately 31,000 polls in total.

Appendix B

Prompt for Poll Labelling

```
content = '''You are an expert at classifying polls into categories.You are also fluent in languages such as \\\
   English, German, French, Spanish.You may be given polls in either of these languages, but you have to categorise it from the \\\
   English categories that will follow in this prompt. The list of categories is in the following list between the square brackets: \\\
   [Work, Vehicles, Games, Books/Reading, Food, Dating, Family, Movies/Television, Friends/Relations, Money/Finance, Health, \\\
         Internet/Technology, Cosmetics/Wellness, Fashion, Music, Politics, Celebrities/Gossip, Travel, School, Shopping, Spirituality, \\\
            Sport, Animals/Nature, Jokes/Humor]. When you categorise a poll, please answer only with the category name you think the poll fits best in. \\\
   You can only answer with one of the categories from the list below. If a poll does not fit into any category, \\\
   please respond only by saying "Other". If you are not sure which category best fits a poll, then respond with "Other" as well. \\\
   If you are confused between which category to choose, or are not sure about the category to choose, please respond with "Other" as well.\\\
   Before responding, you must that check your response meets the following criteria. If it does not, you must amend it before you respond: \\\
   1. Your response should look like this, for example: Work\
   2. If you choose to answer with "Other", please answer by saying: Other \\\
   3. There must be no additional text or sentences other than the category name, or "Other"\
   Each poll will be within square brackets [ ]. Please examine the "
```

```
   question" text and the "answer" options before choosing which category
   the poll fits best in.'''\
 \
def get_completion(prompt, model="gpt-3.5-turbo"):\
  messages=[\
      \{"role": "system", "content": content\},\
    \{\
          "role": "user",\
          "content": json.dumps(\{\
              "text": "hast Du schon einmal etwas von den Projekten Neom
   und Mukaab in Saudi-Arabien geh\'f6rt",\
              "answers": [\
                  \{"text": "Ja"\},\
                  \{"text": "Nein"\},\
                  \{"text": []\},\
                  \{"text": []\}\
              ],\
              "key": [],\
              "rows": []\
          \})\
      \},\
      \{\
          "role": "assistant",\
          "content": "Politics"\
      \},\
      \{\
          "role": "user",\
          "content": json.dumps(\{\
              "text": "Manche Quieckes sind nicht beantwortbar. Bekommt es
    dan ein dislike?",\
              "answers": [\
                  \{"text": "Kommt drauf an"\},\
                  \{"text": "Klar!"\},\
```

```
                \{"text": []\},\
                \{"text": []\}\
            ],\
            "key": [],\
            "rows": []\
        \})\
    \},\
    \{\
        "role": "assistant",\
        "content": "Other"\
    \},\
    \{\
        "role": "user",\
        "content": json.dumps(\{\
            "text": "Was verbindest du phonetisch mit dem griechischen Wort Boufos.",\
            "answers": [\
                \{"text": "Eine Eule"\},\
                \{"text": "Einen unbeweglichen menschlichen Holzklotz"\},\
                \{"text": "Einen klugen Mann"\},\
                \{"text": "Einen hemmungslosen Hedonisten"\}\
            ],\
            "key": [],\
            "rows": []\
         \
        \})\
    \},\
    \{\
        "role": "assistant",\
        "content": "Other"\
    \},\
    \{\
```

```
        "role": "user",\
        "content": json.dumps(prompt)\},\
    ]\
    response = openai.ChatCompletion.create(\
        model=model,\
        messages=messages,\
        temperature=0.1, # this is the degree of randomness of the model's
     output\
    )\
    return response.choices[0].message["content"]}
```

Appendix C

Final dataset preparation

The final dataset was prepared by making a list of users for the analyses on positive implicit interest (analysis 1 and 4) and the analyses on negative implicit interest (analysis 2 and 3). Users who might have upvoted or downvoted a poll would be in a particular dataset (positive or negative implicit interest, respectively). Thus, a user could theoretically feature in both datasets. Therefore, we separated the datasets for positive implicit interests (upvoted polls) and negative implicit interests (downvoted polls), since we would be using them for different analyses.

Positive Implicit Interests dataset

This dataset would be used for analysis 1 and 4, which examine positive and negative explicit interests and positive implicit interests. Thus, it consists of all the users who have upvoted at least one poll. Each row represents one user. The row consists of the user ID (anonymised for data protection), number of polls upvoted for each of the 24 topics (representing positive implicit interest), the total number of polls upvoted by the user, the proportion of upvotes for each topic (for example, the number of upvotes for the topic "food", out of the the total number of upvotes given by the user for all topics), the users' positive explicit interest for each topic (either 1 or 0), and their negative explicit interest (i.e., explicit disinterest) for each topic (either 1 or 0).

Negative Implicit Interests dataset

This dataset would be used for analysis 2 and 3, which examine positive and negative explicit interests and negative implicit interests. Thus, it consists of all the users who have downvoted at least one poll. The structure of the dataset follows that of the positive implicit interests dataset, but looks at downvotes instead of upvotes.

Thus, again, each row represents one user. The row consists of the user ID, number of polls downvoted for each topic (representing negative implicit interest), the total number of polls downvoted by the user, the proportion of downvotes for each topic (for example,

the number of downvotes for the topic “food”, out of the the total number of downvotes given by the user for all topics), and again, the users’ positive explicit interest for each topic (either 1 or 0), and their negative explicit interest (i.e., explicit disinterest) for each topic (either 1 or 0).