I’m Nervous about Sharing This Secret with You: Youtube Influencers Generate Strong Parasocial Interactions by Discussing Personal Issues

Mike Thelwall†, Emma Stuart, Amalia Mas-Bleda, Meiko Makita, Mahshid Abdoli
Statistical Cybermetrics Research Group, University of Wolverhampton, UK

Abstract

Purpose: Performers may generate loyalty partly through eliciting illusory personal connections with their audience, parasocial relationships (PSRs), and individual illusory exchanges, parasocial interactions (PSIs). On social media, semi-PSIs are real but imbalanced exchanges with audiences, including through comments on influencers’ videos, and strong semi-PSIs are those that occur within PSRs. This article introduces and assesses an automatic method to detect videos with strong PSI potential.

Design/methodology/approach: Strong semi-PSIs were hypothesized to occur when commenters used a variant of the pronoun “you”, typically addressing the influencer. Comments on the videos of UK female influencer channels were used to test whether the proportion of you pronoun comments could be an automated indicator of strong PSI potential, and to find factors associating with the strong PSI potential of influencer videos. The highest and lowest strong PSI potential videos for 117 influencers were classified with content analysis for strong PSI potential and evidence of factors that might elicit PSIs.

Findings: The you pronoun proportion was effective at indicating video strong PSI potential, the first automated method to detect any type of PSI. Gazing at the camera, head and shoulders framing, discussing personal issues, and focusing on the influencer associated with higher strong PSI potential for influencer videos. New social media factors found include requesting feedback and discussing the channel itself.

Research limitations: Only one country, genre and social media platform was analysed.

Practical implications: The method can be used to automatically detect YouTube videos with strong PSI potential, helping influencers to monitor their performance.

Originality/value: This is the first automatic method to detect any aspect of PSI or PSR.

Keywords Parasocial interaction; Parasocial relationships; Influencers; YouTube; Social media; semi-PSI

† Corresponding author: Mike Thelwall (E-mail: m.thelwall@wlv.ac.uk; ORCID: 0000-0001-6065-205X).
1 Introduction

Online videos are an important source of information and entertainment for a substantial minority of the world. According to Alexa.com (www.alexa.com/topsites) YouTube was the world’s second most popular website in March 2022, behind Google but ahead of popular video sharing sites including bilibili (China), Youku (China), iQIYI (China), Niconico (Japan), DailyMotion (France), Vimeo (creative content), Twitch (gaming), and video services within Facebook and Instagram. Understanding the strategies used by successful influencers in these sites can inform future content creators and may give insights into the psychological processes underpinning them. This understanding may also be important for marketers engaging with influencers (e.g., Sokolova & Perez, 2021) and for mental health professionals if the influencers impact the wellbeing of their audiences (e.g., de Bérail, Guillon, & Bungener, 2019; Erfani & Abedin, 2018).

The social media video format allows influencers to chat directly to camera in an informal personal setting (e.g., bedroom, living room) and to respond to viewers through video commenting functions. The performers may strive to create a sense of intimacy with viewers to engage their attention (Marôpo, Jorge, & Tomaz, 2020; Raun, 2018). Repeated exposure to performers in naturalistic situations like this can lead audience members to feel a subconscious or conscious social bond with the performer: a parasocial relationship (PSR) (Horton & Wohl, 1956). A PSR is an ongoing feeling of an intimate relationship between the audience and performer (“a lingering sense of intimacy and connectedness with media personalities”, Tukachinsky, Walter, & Saucier, 2020). PSRs build loyalty (e.g., Conway & Rubin, 1991; Rubin & Step, 2000), helping a show or performer to attain popularity, and hence are commercially important. PSRs may also be built partly through individual parasocial interactions (PSIs), instances where audience members perceive an interaction to be occurring (“an illusory give-and-take with media figures”, Tukachinsky, Walter, & Saucier, 2020).

On YouTube, unlike TV and radio, there is a realistic chance of the viewer interacting with the performer through commenting on their video, possibly receiving a Like or reply from the performer or their team. This is like a PSI in that the relationship is between an audience member and performer, with little chance of turning into a personal friendship (e.g., Yuan & Lou, 2020). Nevertheless, it is not a classical PSI because it could be a real interaction, depending on whether the performer chooses to engage. Here are three examples of YouTube comment interactions, paraphrased for privacy: “Long term follower, thanks for always making beautiful, wholesome, helpful, videos packed full of ideas and advice! xx” [reply from influencer: “Wow thank you soooo much 😁”]; “your home is so lovely
and comfortable❤️” [reply from influencer: “Thank you!”]; “Your skin looks so healthy!” [comment liked by influencer]. It is also plausible to believe that an influencer might read comments that they do not Like or reply to. This is an invisible but real two-way interaction (influencer: make video; user: watch and comment; influencer: read comment). Since comments can generate real rather than illusory interactions, it is useful to define this as a new concept, as follows.

- **Semi-PSIs** are attempts to contact a media figure, such as by addressing a fan letter, email, or social media comment to them in the belief that it will be seen. For YouTube, it seems likely that semi-PSIs will be elicited by the same types of videos as PSIs because both involve power imbalanced interactions with media figures, although one is fully illusory. If true, most or all research about PSIs may also apply to semi-PSIs.

Another concept that is useful to introduce is the strength of a PSI or semi-PSI. Since PSIs and semi-PSIs can be separate from PSRs, it seems reasonable to characterize those that are closer to PSRs as stronger. Strength is a key concept because this article introduces a method for assessing it in semi-PSIs.

- The **strength** of a PSI or semi-PSI is the extent to which it engages with a PSR. Whereas a strong PSI or semi-PSI is part of a PSR (e.g., Fatima is looking at me when reading the news because I think we are long-term friends), a weak PSI or semi-PSI is not (e.g., the newsreader is looking at me, how strange). A medium PSI or semi-PSI would be between these extremes (e.g., Fatima is looking at me in a friendly way today). This is different to previous characterizations of PSIs (outside PSRs) as “strong” or “intense” based on the self-reported strength of feeling that an interaction was occurring (Dibble, Hartmann, & Rosaen, 2016; Hartmann & Goldhoorn, 2011). The strength of a semi-PSI in the new sense might sometimes be guessed from the evidence. For example, the comment, “Zoella’s makeup is gorgeous” is less likely to be a strong semi-PSI than the comment, “Your makeup is gorgeous, as always, Zoella! xx”.

It is also useful to distinguish between potential and actual PSI/PSR. This is helpful when discussing the properties of videos or influencers that make them likely to generate PSIs or PSRs.

- The **(strong) PSI potential** of a media offering (e.g., video, TV show, TV series) is its intrinsic likelihood to generate (strong) PSIs with its audience. For example, a video with a presenter chatting informally to camera would have higher strong PSI potential than a sunset video for most audiences. Strong PSI potential will be written sPSI potential. The **PSR potential** of a media figure or series is similarly its likelihood to generate PSRs with its audience.
Most previous PSI/PSR research has investigated either audience characteristics associated with pure PSRs/PSIs or media product characteristics associated with PSI/PSR potential (e.g., looking at camera, informal style, regular episodes).

This article introduces a new method to automatically assess a social media video’s sPSI potential through semi-PSIs in its audience comments, validates the method, and uses it to identify factors associating with sPSI potential in UK female influencers’ YouTube videos. Female YouTube lifestyle influencers seem likely to elicit PSIs because they have large audiences, frequently adopt a direct, personal style and may influence their (presumably young female) audiences (Ladhari, Massa, & Skandrani, 2020; Lee & Lee, 2022; Thelwall & Cash, 2021; Torjesen, 2021). The restriction to the UK is a practical decision to increase homogeneity and reduce the chances of interpretation misunderstandings. The presumably female audiences of female influencers can be expected to have stronger parasocial relationships than the presumably male audiences of male influencers (Tukachinsky, Walter, & Saucier, 2020). The new method to assess a video’s sPSI potential is the proportion of comments containing the term ‘you’. Referring directly to the influencer would be a semi-PSI, as defined above, in contrast to general comments (e.g., “love that colour”, “my nails cost £20”) that may not target the influencer. If validated, the you pronoun proportion semi-PSI indicator would give the first automated method to calculate any type of PSI indicator and, therefore, the first automated method to estimate video sPSI potential. The overarching research questions are as follows.

- RQ1: Does the proportion of comments containing a you pronoun correlate with the sPSI potential of a YouTube influencer video?
- RQ2: Which factors associate with higher sPSI potential for YouTube influencer videos?

2 Background

As introduced above, this study adopts and adapts the existing theory of parasocial relationships to video-based social media influencers. The study makes the assumption that existing PSR/PSI theory, although developed for radio and extensively applied to television, also applies to video-based social media influencers. It extends this theory by introducing the concept of semi-PSI and hypothesising that semi-PSI works similarly to traditional PSI. To be clear, semi-PSI is hypothesised to be additional to PSI rather than the social media version of it. PSIs and semi-PSIs are assumed to have different strengths according to the extent to which they reflect underlying PSRs. This study also extends PSR/PSI theory by introducing the concept of PSI/PSR potential for influencers or products, which has been implicit in prior research. Finally, at a methodological level, it introduces you pronoun proportions with the hypothesis (explicitly tested as H1) that they may be indicators of sPSI potential.
2.1 PSRs, PSIs, semi-PSIs, and you pronoun proportions in social media

YouTube influencers create an emotionally safe space (Mardon, Molesworth, & Grigore, 2018), usually talking informally to the camera as if to a friend, which promotes PSR formation (van Reijmersdal et al., 2020). PSR strength strongly correlates (Pearson: 0.6) with motivation to watch knitting vlogs (Gregg, 2020) and associates with time spent watching YouTube by children aged 9–12 (Tolbert & Drogos, 2019). Thus, PSRs seem to be effective for retaining YouTube viewers. They can also generate advertising and marketing revenue for the influencer through enhanced endorsement credibility (Dekavalla, 2020; Ladhari, Massa, & Skandrani, 2020; Sokolova & Kefi, 2020; Reinikainen et al., 2020; Trivedi & Sama, 2020) and make their public service messages more persuasive (Rosaen, Dibble, & Hartmann, 2019; Sakib, Zo Iffagharian, & Yazdanparast, 2020).

Semi-PSIs are also important on social media. Audiences may “Like” a video or comment underneath it, perhaps believing that the influencer will notice or respond (Reinikainen et al., 2020). The performer, in turn, relies on evidence of interaction to be visibly successful (Chen, 2016) and to promote their work on YouTube (Bishop, 2018). This incentivises them to interact and encourage audience responses (Tur-Viñes & Castelló-Martínez, 2019).

There is limited evidence that commenting (i.e., semi-PSIs) can be an indicator of PSRs or strong PSIs. One study found a statistically significant weak positive relationship between commenting on videos and PSR strength for a snowball sample of German-speaking people following any YouTube celebrity (Rihl & Wegener, 2019). For PSIs, viewers told to comment on and like/dislike a video in an experiment (i.e., semi-PSIs) felt stronger PSRs (Pearson: 0.6) with a US female travel vlogger promoting a product (Munnukka et al., 2019). The current study for the first time singles out comments that seem to indicate stronger semi-PSI: those that address the performer directly with you pronouns. The underlying belief is that these apparently stronger semi-PSIs are better indicators of PSI than comments alone. This generates the first hypothesis, which is leveraged for a method to address subsequent hypotheses and analyses.

- **H1 You pronouns:** videos with higher proportions of comments containing variants of the pronoun ‘you’ have higher sPSI potential.

2.2 Factors associated with increased sPSI/PSR potential

The terms PSR and PSI have been used interchangeably in some parasocial literature (Dibble, Hartmann, & Rosaen, 2016) and more is known about PSRs than PSIs, so this section mainly discusses PSRs. The two are definitionally distinct (Dibble, Hartmann, & Rosaen, 2016; Hartmann & Goldhoorn, 2011) but likely to
be connected in practice because repeated PSIs could help build a PSR, a PSR seems likely to make PSIs more common, and the two concepts correlate moderately (Tukachinsky, Walter, & Saucier, 2020). Whilst the correlation and possible connections do not prove that factors affecting PSRs also affect PSIs, factors affecting PSRs at least form a plausible starting point for investigating factors affecting PSIs. Moreover, the current article focuses on strong PSI, which has a definitional connection with PSR. The rest of this section primarily introduces hypotheses about the sPSI potential of videos through prior PSI or PSR research (sometimes also from research conflating the two or implicitly focusing on sPSI). As explained in the methods, these hypotheses will be tested through semi-PSI evidence from you pronoun proportions. Known PSR factors that cannot feasibly be related to the data collected are not included (e.g., Performer physical and social attractiveness [needs viewers’ subjective judgements]; binge watching and episode regularity [no viewing data]). Four mimic informal communication between friends and one is specific to PSRs.

**Gaze:** When friends talk it is normal to look at each other, so gaze is a logical source of PSIs. This has been validated experimentally by showing that a recorded person gazing directly at the camera was more likely to elicit PSIs than when looking elsewhere (Dibble, Hartmann, & Rosaen, 2016; Hartmann & Goldhoorn, 2011; Oliver et al., 2019). This has not been tested on YouTube, but looking at the camera and being closer to it associates with more viewers (Biel & Gatica-Perez, 2011), perhaps by generating enjoyment with PSIs or loyalty with PSRs.

- **H2 Gaze:** An influencer talking primarily to camera in close up (head and upper body) has higher sPSI potential.

**Realism:** Friendships are part of normal life whereas entertainment can take place in artificial circumstances, so PSRs may form more intuitively in more naturalistic media settings that mimic everyday life. There is wide evidence that the perceived realism of a performance influences PSR strength (Giles, 2002; Rubin & Perse, 1987). Similarly, inauthentic brand endorsements in commercials may harm performers’ PSRs (Alperstein, 1991). On social media, realism was related to gaze in one study: YouTube videos where the performer’s face was not visible were judged less authentic (Ferchaud, Grzeslo, Orme, & LaGroue, 2018). On YouTube, lifestyle influencers typically mimic informal chats between friends, so realism equates with a friendly approach.

- **H3 Realism:** A video perceived to be a realistic portrayal of the influencer as a friend has higher sPSI potential.

**Self-disclosure:** Self-disclosure can vary from simple personal details, such as the performer’s real name, to information of the kind that would normally only be revealed to friends, such as relationship problems, bullying, mental health issues,
physical health deterioration, and fears about the future. Self-disclosure is part of forming close friendships and can therefore help performers build PSRs (Giles, 2002). On social media, the (perceived) self-disclosure of beauty vloggers associates weakly (Pearson 0.2) but statistically significantly with PSI strength, according to a convenience sample survey of young college-educated females (Ko & Wu, 2017). Positive and negative self-disclosure associate with realism in videos from the top ten YouTubers (Ferchaud et al., 2018), giving another reason why they may generate PSIs. Since self-disclosure is core to the lifestyle YouTube channel genre, the focus here needs to be on the depth of the self-disclosure rather than its presence. As a practical step, this is equated with novel personal information sharing.

- **H4** Self-disclosure: A video discussing personal issues that the influencer has previously not shared in public has higher sPSI potential.

**Time:** Friendships take time to build through repeated interactions. Parasocial relationships have been hypothesized to form similarly in four stages, as follows. During initiation, the viewer gets first impressions of the performer, singling them out from the many performers they see. In the experimental stage, they find out a broad range of general information about the performer. During intensification, the viewer discovers more personal information about the performer. Finally, the integration stage involves accepting the relationship as part of the viewer’s identity, such as through identifying publicly as a fan (Tukachinsky & Stever, 2019; Giles, 2002). Thus, viewers of later videos may be more likely to have formed a PSR and engage in sPSIs rather than PSIs, if viewers watch chronologically or the proportion of first-time video watchers does not increase over time.

- **H5** Time: An influencer’s more recent videos have higher sPSI potential.

**Popularity:** The popularity of the performer is known to elicit PSRs even though it may not mimic interpersonal friendships. Since PSRs generate loyalty, it is logical to expect performers generating more PSRs to have bigger audiences. On social media, PSRs are stronger for German language vloggers with more viewers (Rihl & Wegener, 2019). Similarly, more popular beauty vloggers are more likely to elicit PSRs with their viewers (Rasmussen, 2018). Given that social media popularity associates with PSRs, it seems likely that is will also tend to have a greater share of sPSIs compared to weak or all PSIs.

- **H6** Popularity: Influencers with more subscribers (i.e., more popular in one sense) produce videos with a higher proportion of sPSIs compared to PSIs.

**Unknown factors:** This study will also seek factors suggested by the data that have not been found in the literature review, as a final qualitative objective.

- **Qual1:** Which types of videos have higher sPSI potential than suggested by the above factors?
3 Methods

The research design was to obtain a large sample of popular UK female YouTube lifestyle influencers and obtain statistical evidence of their popularity and the popularity of their videos from YouTube, in addition to you pronoun proportion data for each for their videos. A paired design was chosen, comparing two videos from the same channel rather than comparing between channels because each channel has a different style and focus (Figure 1). The paired design also partly factors out the average audience PSR strength for individual influencers by contrasting two videos from the same influencer. Comparing videos between different types of influencer could have therefore introduced spurious relationships from the different topics covered. The research design was as follows, and is described and justified in detail below.

- H1: test whether videos judged to have higher sPSI potential tend to have higher you pronoun proportions.
- H2, H3, H4: test whether videos with higher you pronoun proportions (or higher judged sPSI potential) are more likely to use gaze, friendliness and self-disclosure (if H1 is accepted).
- H5: test whether newer videos from an influencer tend to have higher you pronoun proportions (if H1 is accepted).

Figure 1. The manual and automated stages of the data collection and classification components of the research design. Results from earlier stages are also used in some statistical analyses (e.g., for popularity).
• H6: test whether more popular influencers’ videos tend to have higher you pronoun proportions (if H1 is accepted).
• Qual1: manually examine videos with high you pronoun proportions despite not having high scores on H2, H3, and H4 for possible additional video sPSI potential factors (if H1 is accepted).

As described and justified below, the two videos chosen for each influencer had the highest and lowest proportion of comments containing the word you. The research design is mixed methods, with quantitative approaches for some hypotheses and qualitative methods or both for others. A mixed methods design was necessary because H1 is primarily quantitative whilst H2, H3, H4, and Qual1 have qualitative components. In particular, human judgement is needed to identify sPSI potential factors within videos. Various forms of statistical analysis were applied to the information gathered about the pairs of videos, and indirect heuristic-based analyses were used to investigate hypotheses that could not be addressed with the paired design. In particular, if the you pronoun proportion is validated as a sPSI potential indicator, it can be used for the time analysis when there are too many videos to be human coded.

3.1 Female YouTube influencer channel selection

There is no register of lifestyle influencers on YouTube, so a large set was identified by browsing YouTube and the wider web. Sources included the Lifestyle YouTube category, recommendations in video comments, online magazine articles about prominent influencers, subscription/following lists on YouTube and Instagram for people with lifestyle interests, and searches in YouTube (e.g., “make-up haul” “fashion haul”). YouTube was difficult to search for a comprehensive set of relevant influencers, so all authors searched and browsed separately using ad-hoc queries, browsing and intuitive explorations to make a reasonably comprehensive list, although the final list is almost certainly incomplete. Lifestyle influencers’ channels were retained if they matched the following criteria:

• A single female influencer or (in three cases) two female influencers working as a team. For consistency, female influencers working with male partners were excluded, but those occasionally making videos with a male partner were allowed. Trans women would have been included, of course, had any been found.
• British and living in the UK. An influencer was assumed to be British if they had a British accent and declared a UK location or had a UK focus.
• A focus on fashion or lifestyle but not cooking since cookery vlogs have a narrower focus and a “how to” style. Fashion and lifestyle content seemed to have a large overlap, so no attempt was made to distinguish between the two.
• At least 2,000 subscribers at the time of data collection so that they could be regarded as broadly popular (for reference, the first author’s six year-old channel with 30 videos had only 54 subscribers by June 2021).

The influencers were from different ethnicities, apparent social class backgrounds, sexualities and ages, although most seemed to be aged 19 to 35. Many channels had unique characteristics, including a focus on family, make-up, cruelty-free make-up, luxury fashion, high street fashion, student life, travel, and fitness. Some influencers had evolved through romantic partnerships, marriage, separation, and children. They may also occasionally address issues that do not normally fit their theme or create spin-off channels for different content.

The final set included 289 UK female influencer channels, 286 individual females and 3 female duos. Two channels were from the same person (Zoella and Zoe Sugg), which were retained since Zoella is a highly successful influencer and these channels contain differing content types. A second pair of channels (Rose and Rosie, Rose and Rosie Vlogs) were from the same team but with a different style and attracting a million subscribers for one and nearly half a million for the other, so these were also retained. On 6 December 2020 these channels collectively accounted for 111,194,290 subscribers (including duplicates, excluding one channel hiding its subscriber count; median 107,500 per influencer) and 11,811,612,486 individual video views (median 9,059,454 per influencer). Only registered YouTube users can comment or subscribe, so at least a median of 107,000 viewers per influencer were enabled to comment.

3.2 Comment data

YouTube has an Applications Programming Interface (API) that allows lists of videos to be requested for any channel as well as the comments on each video. This service (version 3) was used in the software Mozdeh (https://mozdeh.wlv.ac.uk) to download all the comments on all the videos of all the influencers in the set. The data collection took place between 29 October and 2 December 2020 from computers in the UK. A user can click reply to respond to a comment by another user and these replies are flagged as such by the information in the API data. All replies were excluded because the focus was on parasocial interaction with the influencer rather than interactions between commenters. This stage produced 21,910,526 comments on 76,923 videos from 289 UK female lifestyle influencer channels.

3.3 Comment pronoun data

The comments were processed to identify semi-PSI evidence. Comments were classified as semi-PSI, likely addressing the influencer, when they contained a variant
of the pronoun “you” (you, you’ve, you’re, your, yourself, yourselves). Since replies to other commenters were excluded, a comment written underneath the video of an influencer containing a you pronoun seems likely to address the influencer. To test this, 100 matching comments were selected with a random number generator, and all appeared to be addressing the influencer directly. In contrast, only 7% of 100 non-matching comments directly addressed the influencer (e.g., with “u”, “ur”, or their name) and a further 11% addressed the influencer through a request (e.g., “Do lips next”) or implicitly (e.g., “happy birthday”, “thanks”, “congratulations”, “keep the black dress”). Thus, you pronouns in comments seem to be more frequently attempts to interact with the influencer than other comments, on average.

For each video, the proportion of (non-reply) comments containing a variant of the you pronoun was calculated using a program written for this (available as a button in the YouTube tab of the free software Webometric Analyst at https://lexiurl.wlv.ac.uk). The ‘you’ proportion statistic is imperfect because it does not take into account other reasons for commenting. In an extreme case, for example, a video might elicit much semi-PSI but a low you proportion because even more comments responded to a comment-based competition or other request in the video. H1 (methods described below) assesses the pronoun method as a validity check, by correlating it with manually assessed video sPSI potential.

Videos with fewer than 100 comments were excluded because the you proportion would be a less reliable semi-PSI indicator for these. The 100-comment threshold was chosen in advance as a simple round number, allowing proportions to be accurate to 1%. This was chosen as a trade-off between a larger threshold with fewer videos and a lower threshold with less accurate proportions. After excluding videos with fewer than 100 comments, influencers with fewer than 100 remaining videos were also excluded, leaving 118. The arbitrary threshold of 100 was set to increase the likelihood that there was substantial parasocial variation within the set for an influencer. On 6 December 2020 these 118 channels collectively accounted for 91,001,600 subscribers (including duplicates, excluding one channel hiding its subscriber count; median 437,000 per influencer) and 10,533,276,527 individual video views (median 45,035,114 video views per influencer).

### 3.4 Paired videos manual classification

A manual content analysis of videos was used to identify PSI factors (for H2, H3, H4) and overall sPSI potential (for H1). For this, a sample of videos was selected to represent likely high and low sPSI potential. For each influencer, the video with the highest you pronoun proportion (henceforth: high you video) and the video with the lowest you pronoun proportion (henceforth: low you video) were selected for manual investigations of factors from H2 to H4. These represent apparent sPSI
extremes for each influencer and this sample should therefore be easiest to analyse for sPSI potential differences. A paired design (one high and one low you video per user) is more robust than selecting the overall highest and lowest you videos because the latter approach might primarily reflect the strategies of a single influencer. Moreover, a single influencer’s strategy might relate to sPSI in an unusual way.

The paired design resulted in 236 videos, two from each of the 118 influencers with at least 100 videos having at least 100 comments each. During the manual classification, one of the videos became private and the channel was excluded, leaving a final set of 117 channels and 234 videos.

The paired videos were manually classified for the presence of factors and overall sPSI potential (based on the Wikipedia description of PSI which describes sPSI). The former was used to assess the hypothesised video-based sPSI factors (H2, H3, H4) and the latter to validate the methods assumption that the proportion of you pronoun comments is an indicator of sPSI potential for YouTube videos (H1).

Four researchers with experience of coding social media sites conducted the content analysis. A codebook was created to describe the classes to be used, based on the research hypotheses and intuitions obtained by browsing video channels before the data collection started. Because YouTube influencers often seem to follow a direct camera gaze style, H2 was split into five related parts to investigate it at a fine-grained level. Some facets were coded by watching the first 3 minutes and the last 2 minutes (5 minutes in total) and other facets by also hovering the cursor along the rest of the video:

- Watching 5 minutes: Eye contact, Main focus, Friendly, Personal issues disclosed, High sPSI potential.
- Watching 5 minutes and cursor hovering: Faces camera, Head and shoulders, Alone.

For the pilot testing phase, 134 videos were selected from the 76 influencers that had at least 50 but fewer than 100 videos with at least 100 comments each to train the coders and fine-tune the codebook. Thus, there was no overlap in the influencers between the pilot set and main set. To help with training, the pilot sample included the classification of videos as high or low you (based on you proportions, but the coders were not told about the you pronoun heuristic used) and the main sample did not. The coders were instructed not to read the comments under a video but make classifications based on the video alone. One of the aspects coded was the sPSI potential of the video. Although PSI and PSR are normally investigated using validated instruments asking multiple questions, a direct question was more appropriate because the coders were not interacting with the videos as viewers and understood the PSI concept, relating it to PSR. Inter-coder agreement was assessed
on the pilot sample, discussing cases of disagreement. The same people then coded
the main sample, with two coders independently classifying each video, and each
coder paired with the other coders for approximately an equal number of videos
(i.e., a partial crossed design).

Krippendorff’s interval alpha was used to assess the levels of agreement between
coders on the six classes that reference percentages of the videos with the given
feature, and ordinal alpha for the two classes using qualitative scales. Krippendorff
recommends a score of at least 0.8 for firm conclusions and 0.667 for tentative
conclusions (Krippendorff, 2019) when the results are used for content analysis
purposes. The inter-coder agreement is adequate for tentative conclusions, except
for the Friendly class (Table 1). For completeness, the Friendly class results are
reported in the paper alongside the others, but these are suggestive rather than
sufficient for tentative conclusions.

Table 1. Inter-coder consistency scores for facets of UK female influencer videos Krippendorff’s alpha at the
difference between high you and low you videos within a channel level (n=117). Interval alpha is used for the
middle six and ordinal alpha for the other two.

<table>
<thead>
<tr>
<th>Facet</th>
<th>1</th>
<th>2,3,4</th>
<th>5</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 High sPSI potential</td>
<td>None</td>
<td>....***</td>
<td>Very high</td>
<td>0.716</td>
</tr>
<tr>
<td>H2 Faces camera</td>
<td>Almost never</td>
<td>....*</td>
<td>Virtually all the time</td>
<td>0.730</td>
</tr>
<tr>
<td>H2 Eye contact with camera lens</td>
<td>Almost never</td>
<td>....*</td>
<td>Virtually all the time</td>
<td>0.684</td>
</tr>
<tr>
<td>H2 Head and shoulders mainly in shot</td>
<td>Almost never</td>
<td>....*</td>
<td>Virtually all the time</td>
<td>0.722</td>
</tr>
<tr>
<td>H2 Alone (no other talking people)</td>
<td>Almost never</td>
<td>....*</td>
<td>Virtually all the time</td>
<td>0.936</td>
</tr>
<tr>
<td>H2 Main focus is vlogger</td>
<td>Almost never</td>
<td>....*</td>
<td>Virtually all the time</td>
<td>0.842</td>
</tr>
<tr>
<td>H3 Friendly, realistic chat style</td>
<td>Almost never</td>
<td>....*</td>
<td>Virtually all the time</td>
<td>0.552</td>
</tr>
<tr>
<td>H4 Personal issues disclosed</td>
<td>Never</td>
<td>....**</td>
<td>Very personal focus</td>
<td>0.887</td>
</tr>
<tr>
<td>Qual1 Reason for high or low PSI</td>
<td>Free text</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Almost never - 25% of the time - half of the time - 75% of the time - Virtually all the time.
** Never - Some personal but not private issues discussed - A lot of personal but not private issues discussed -
Some previously private issues discussed, but not the focus of the video -The video focuses on very personal
and previously private issues.
*** Unlikely to engage the viewer with parasocial interaction in any of the video - Unlikely to engage the
viewer with parasocial interaction in most of the video - Moderately likely to engage the viewer with parasocial
interaction in most of the video - Likely to engage the viewer with parasocial interaction in most of the video
- Very likely to engage the viewer with parasocial interaction in most of the video.

3.5 Pronoun heuristic assessment (H1)

The effectiveness of the you pronoun proportion as an indicator of sPSI potential
(H1) was assessed by calculating a 95% confidence interval for the average
difference between the sPSI potential of each high you videos subtract the sPSI
potential of the corresponding low you video. A confidence interval of positive
values excluding zero would indicate that the you pronoun proportion is a reasonable
method to assess sPSI potential for individual (successful UK female) influencer
videos. The limitation of this test is that it only considers extreme cases (the highest
3.6 Analysis of video-based PSI factors (H2, H3, H4)

Spearman correlations were calculated between the sPSI potential difference (high you video score minus low you video score) and PSI factor score differences (high you video score minus low you video score) for each channel. This test assesses whether the three hypothesised PSI factors tend to occur more in videos judged to have higher sPSI potential. This test may have a confirmation bias because the same coders judged all factors so they might be more likely to judge a video as having sPSI potential if they found factors known to associate with it.

To avoid confirmation bias, 95% confidence intervals were calculated for the average difference between the high and low you videos for each factor. If these confidence intervals are positive and exclude 0 then this gives statistical evidence that the factors associate with high you pronoun proportion videos. Combining this with a positive result for H1 would give evidence, untainted by confirmation bias, that the factors associate with higher sPSI potential.

3.7 Time series analysis (H5)

The 118 influencers with at least 100 videos containing at least 100 comments were also investigated to assess whether the you proportion, as an indicator of sPSI potential, changed over time. For each influencer, their videos with at least 100 comments were arranged in order of publication date and a Spearman correlation calculated to assess whether there was a tendency for the you proportion to increase or decrease over time. The average of the Spearman correlations is an indicator of the average tendency across channels.

3.8 Popularity vs PSI (H6)

Spearman correlations were calculated to assess whether more popular influencers’ videos tended to elicit more strong semi-PSI, as reflected by the median proportion of you-related pronouns in comments on their videos. This test was carried out for all 118 channels in the main dataset, including the channel that was excluded from the content analysis for its missing video. The median you proportion across all videos was used to reflect the overall channel average. This test is dependent on H1 validating you pronoun proportions as an indicator of sPSI potential.

3.9 Ethics

This study uses only public data and thus was exempt from ethical approval. To protect the privacy of commenters by not drawing attention to them, no commenter...
names or exact quotes are given (Wilkinson & Thelwall, 2011). Influencers are only named if they are prominent media figures for the same reason. The four influencers named in this article all have over a million subscribers, which seems sufficient to classify them as public figures.

4 Results

4.1 You proportions as a sPSI potential indicator (H1)

Videos with the highest you proportion tended to have higher sPSI potential, as judged by the coders (Figure 2), with the difference being on average 1.18 (95% CI: 1.02 to 1.35; sd=.90). Moreover, in 98 out of the 117 channels, the coders judged the high you video to have higher sPSI potential than the low you video, with only 10 reverse cases. Thus, you proportions are reasonably effective at differentiating between high and low sPSI potential videos. As a corollary, you proportions can be used as an indicator of sPSI potential (because they correlate) and it seems reasonable to use them also as indicators of stronger semi-PSI, as defined above (because the correlate with the related sPSI potential concept, and are calculated from comment interactions).

Figure 2. The difference between the sPSI potential of the highest you video and the lowest you video by channel (e.g., a channel would score 3 if the highest you video had sPSI potential 5 and the lowest you video had sPSI potential 2), in the 117 channels analysed.

The ten cases where the video with a lower you proportion had a higher sPSI potential score were examined for possible causes, as an extra validation step. In one case, a low PSI video included a request by the influencer to comment with a specific phrase to enter a competition. This phrase contained “you”, which inflated the you proportion in a non-sPSI way. In three cases the high you proportion video
with low sPSI potential discussed the channel itself, with commenters making suggestions. This seems to be an example of genuine interactions. Three other videos had small segments discussing personal issues upsetting the influencer, although the focus of the video was something else, suggesting that even brief mentions of personal issues can generate sPSIs. The reason for the discrepancy in the other two videos was unclear.

4.2 PSI factors and sPSI potential (H2, H3, H4)

Based on the average coder scores, within each channel, videos with higher sPSI potential were more likely to have stronger degrees of all individual factors classified: influencer facing the camera, making eye contact with the camera, shooting mainly their head and shoulders, presenting alone, focusing on themselves, talking in a friendly and realistic style, and mentioning personal issues (Table 2). The Spearman correlations supporting this conclusion are for differences between the score in each dimension for the high and low you videos for each channel. The reason for the weaker associations between some factors and sPSI potential may be that the characteristic was either nearly universal or rare. The influencer was almost always facing the camera in almost all videos, for example, and was also usually alone (Table 3). In this context, the weakest association may be for the head and shoulders shot style, since this is much less universal (3.4 instead of 4.4) but has only a slightly stronger correlation with sPSI potential (0.262 instead of 0.224).

Recalling that the correlation results above may be influenced by confirmation bias since sPSI potential was judged by the same people that judged the other factors, comparing between high and low you videos avoids this possibility, however, and has been validated by H1 being accepted. The 95% confidence intervals for the difference between high and low you videos for each factor are all positive and exclude 0, confirming the correlation results above (Figure 3).
As mentioned above, low values for the confidence intervals can occur either because a factor has little effect or because there is little variation in the factor. An example of a factor with little variation is Faces camera (Figure 4), for which the high and low you videos in 45 of the 117 channels had the same score. Taking this into account, in all cases it is rare for channels to have higher scores on the low you video than on the high you video for any factor (Figure 4).

### 4.3 PSI exception videos (Qual1)

A qualitative examination of the high you videos that did not have the typical sPSI potential factors suggested that content focusing on the channel itself or requesting viewer feedback may generate strong semi-PSIs, or at least a high proportion of you pronoun comments. The exceptions included videos with brief introductions to personal issues, suggesting that even this small amount can be effective at generating strong semi-PSIs.
4.4 sPSI potential of videos from a single influencer changing over time (H5)

The proportion of comments containing “you” was correlated against publication date for channels with at least 100 videos with at least 100 comments each (n=118). The median Spearman correlation was -0.033, indicating a slight tendency for semi-PSI, as indicated by you proportions, to decrease over time. This tendency occurred for a small majority of influencers (56%), although this is not statistically significantly different from 50% (Binomial test). Thus, there is insufficient evidence to conclude that the sPSI potential of videos from an influencer tends to increase or decrease over time (publication date).

The two thresholds were relatively arbitrary, so alternative values were tried as a robustness check. Varying the two thresholds (100, 50, or 10 comments, 100, 50, or 10 videos) had little effect on the Spearman correlation (always between -0.02 and -0.05) but the median correlation was statistically significantly different from 0 in two of the nine cases examined (Table 4). Thus, it is possible that there would be statistical evidence for a tendency for sPSI potential to decrease over time within a UK female influencer channel for some conceptualizations of the problem, but the correlations seem too small to be of practical interest.
I’m Nervous about Sharing This Secret with You: Youtube Influencers Generate Strong Parasocial Interactions by Discussing Personal Issues

Table 4. Median Spearman correlation between video age and you pronoun proportion within a channel for different thresholds.

<table>
<thead>
<tr>
<th>Spearman’s rho</th>
<th>≥ 10 videos</th>
<th>≥ 50 videos</th>
<th>≥ 100 videos</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 100 comments per video</td>
<td>-0.033 (n=118)</td>
<td>-0.041 (n=148)</td>
<td>-0.021 (n=194)</td>
</tr>
<tr>
<td>≥ 50 comments per video</td>
<td>-0.037* (n=151)</td>
<td>-0.028 (n=185)</td>
<td>-0.030 (n=236)</td>
</tr>
<tr>
<td>≥ 10 comments per video</td>
<td>-0.049** (n=211)</td>
<td>-0.029 (n=247)</td>
<td>-0.021 (n=275)</td>
</tr>
</tbody>
</table>

*Rho significant at the 0.05 level (2-tailed). *Rho significant at the 0.01 level (2-tailed).

4.5 Popularity and sPSI potential factors (H6)

There was a statistically significant weak tendency for more popular influencers to attract less strong semi-PSI, at least in the form of you-related pronoun proportions (Table 5). This tendency held at all three thresholds assessed.

Table 5. Spearman correlations between the median proportion of you-related pronouns in comments on videos and the total number of subscribers or video views, for different minimum numbers of qualifying videos. One channel with hidden channel subscription counts was excluded.

<table>
<thead>
<tr>
<th>Qualification</th>
<th>Total channel views</th>
<th>Channel subscribers</th>
<th>No. of channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 100 videos with ≥ 100 comments</td>
<td>-0.260**</td>
<td>-0.229*</td>
<td>117</td>
</tr>
<tr>
<td>≥ 50 videos with ≥ 100 comments</td>
<td>-0.216**</td>
<td>-0.193*</td>
<td>147</td>
</tr>
<tr>
<td>≥ 10 videos with ≥ 100 comments</td>
<td>-0.252**</td>
<td>-0.250**</td>
<td>192</td>
</tr>
</tbody>
</table>

* Significant at the 0.05 level (2-tailed).
** Significant at the 0.01 level (2-tailed).

5 Discussion

A limitation of this study is its focus on individual PSIs of various types rather than longer-term PSRs, which may be more important. Although previous survey-based studies have found a moderate positive correlation between the two (Tukachinsky, Walter, & Saucier, 2020), it should not be assumed that the same is true for the types of PSI examined here. It also does not consider relationship building over time (Tukachinsky & Stever, 2019) or other channels used by the influencer, such as Instagram, which may help to build or to deepen PSRs (Giles, 2002; Tukachinsky & Stever, 2019). In addition, a few comments are not in English and may contain pronouns in other languages. Some uses of the you pronoun do not refer to the influencer, and there are other ways of addressing them (e.g., by name), so the data is incomplete.

5.1 You proportions as a sPSI potential indicator (H1)

The results support H1, providing evidence that you pronoun proportions in comments is a valid sPSI indicator for UK female YouTube influencer videos. The statistical tests add to the analytic argument above that referring to an influencer as ‘you’ indicates a stronger type of interaction than usual semi-PSI YouTube comments.
This is the first study to investigate this indicator. It seems likely that the you pronoun proportion indicator would be valid for female lifestyle influencers in English-speaking nations and similar indicators may work in other languages. The extent to which this indicator works for other types of influencer or video content is unclear, however.

The you pronoun indicator seems likely to reflect the four stages of PSR formation to different extents, with stronger associations with later stages. This is based on the hypothesis that the stronger the PSR, the more likely the viewer would be to comment directly to the performer. Some “you” comments were negative, criticising the influencer for boring content. These could be evidence of PSR breakdown (i.e., late stage PSR: Tukachinsky & Stever, 2019) from previously satisfied viewers or evidence of blocked PSR from unimpressed first-time viewers. These comments seemed to be a small minority, so they were not singled out for removal.

### 5.2 sPSI factors and sPSI potential (H2, H3, H4)

The results confirm that three factors known to elicit PSIs or PSRs in other contexts also apply to female YouTube influencers: talking directly to the camera (assessed in different ways) (Dibble, Hartmann, & Rosaen, 2016; Hartmann & Goldhoorn, 2011; Oliver et al., 2019), revealing private and personal information (Giles, 2002; Ko & Wu, 2017) and (although not validated) using an informal friendly style (Giles, 2002; Ru; Tukachinsky, Walter, & Saucier, 2020). Of these, revealing private and personal information seems to be the strongest marker of sPSI potential, as found both by the ‘you’ proportion confidence intervals and the classification correlations. This is partly because influencers seem to vary this factor more than the others examined: not all videos can reveal new personal issues. Personal issues were sometimes discussed briefly at the start of a video, before switching to the usual style. Two of the coders thought that the emotionality was false in some videos, but it is not clear whether typical viewers would agree.

Influencers displayed emotions when discussing personal issues. These emotions sometimes manifested in slight changes in tone of voice and expression (as is normal: Van den Stock, Righart, & de Gelder, 2007) but also sometimes in (appropriate) tearfulness. It has previously been remarked that crying influencers can create intimacy with their audiences and perhaps also authenticity (Berryman & Kavka, 2018). It is therefore not clear whether the personal information or the emotions elicited PSIs, although a combination of the two may be the most powerful and most authentic.

### 5.3 sPSI potential of videos changing over time (H5)

The results suggest no sPSI potential change over time (publication date) for the videos of an influencer, or a small decrease over time. PSRs should strengthen over
time for individual audience members (Tukachinsky & Stever, 2019), so it might be expected that sPSIs would also be more common for later videos. Nevertheless, even if videos are mainly watched when they are recent and are watched chronologically, time series comment statistics partly reflect loyal repeat visitors and partly reflect new visitors. Thus, it is possible that sPSIs are more common for successful channels for repeat visitors, but that this is counterbalanced by the increased number of new visitors hearing about the channel. The results are therefore not directly comparable to prior research.

5.4 Popularity and sPSI factors (H6)

The negative relationship between influencer popularity and the you pronoun sPSI potential indicator suggests that more popular influencers do not produce videos with higher sPSI potential. Whilst this has not been tested before in prior research, it is unexpected because more popular influencers elicit stronger PSRs (Rasmussen, 2018; Rühl & Wegener, 2019). A possible explanation is that you pronoun commenting is less likely for popular influencers because their viewers would face more competition for attention from the larger audience. Thus, more popular influencers may generate stronger PSRs and more sPSIs but weaker semi-PSIs. Alternatively, the fame of popular influencers may attract a larger share of new viewers, with these interacting less than long term viewers with PSRs.

6 Conclusions

This study has made the following theoretical, methodological, and empirical contributions to the literature. They are briefly contextualised in terms of prior knowledge here, and their practical implications are suggested.

- Introduced three new concepts that were implicit, overlooked or underdeveloped in prior research:
  - The sPSI/PSI/PSR potential of media offerings or performers. This has been implicit in prior research but seems like useful general terminology as well as a concept that may help performers to more easily consider this aspect of their offerings.
  - The strength (in the sense of being part of a PSR) of PSIs (including the concept of a strong PSI, sPSI). Prior research has discussed the strength of feeling of PSIs divorced from PSRs (e.g., in experimental conditions) but on social media evidence of PSIs can be harvested from the web, so it is useful to conceptually differentiate between fleeting PSIs and those that suggest a PSR. Since PSRs are important for performers to engage and retain audiences, methods to identify them are useful.
Semi-PSIs as different from PSIs. Prior research has arguably overlooked the difference between imaginary and real interactions when using the term PSI. Introducing semi-PSI clarifies that the two are distinct, suggests that they might be related and may avoid future confusion about the status of social media comments in particular.

- Argued that directly addressing an influencer in comments is more suggestive of a PSR (i.e., more likely to be a strong PSI) than comments that do not. This sets up the automated method introduced below but influencers could also use this as a simple check of the engagement of their audience. For example, if they notice that comments on a video rarely address them directly (by name or with a you pronoun) then this would flag that the content may have made a weak personal connection.

- Introduced the first automated method to detect anything related to PSI or PSR: the you pronoun proportion for detecting the sPSI potential of YouTube videos. The existence of an automated method opens the door to much larger scale research than possible before (including the current article) and new applications that use sPSI potential for other purposes.

- Validated the you pronoun proportion method on female UK lifestyle influencers, an influential group with over 12 billion views for the videos in this article. Whilst the method has been validated on one specific (albeit large) group, it seems likely that it may apply to a wide range of social media genres that allow commenting. The method would presumably be most useful when applied to multiple videos from the same genre (e.g., all home maintenance vloggers) or to the videos of a single influencer to find those with the highest and lowest sPSI potential.

- Confirmed that three factors known to influence PSIs or PSRs also apply to female UK lifestyle influencers (gaze, realism, self-disclosure). This is a relatively minor contribution, confirming a result that would be strongly expected. Perhaps the most interesting aspect of the result is that deeper self-disclosure is particularly important for these influencers, despite the entire genre being based on self-disclosure. To illustrate this point, whilst a car vlogger might occasionally self-disclose by introducing her home or wife, such revelations are normal for lifestyle influencers, who can nevertheless “escalate” self-disclosure, such as by discussing previously suppressed childhood bullying traumas.

- Found that one factor known to influence PSI does not apply to female UK lifestyle influencers: popularity. Because this finding is unexpected, it needs further research with method triangulation to assess whether it reflects a different way that YouTube lifestyle influencing works or whether it is an
I’m Nervous about Sharing This Secret with You: Youtube Influencers Generate Strong Parasocial Interactions by Discussing Personal Issues

Mike Thelwall et al.

Research Paper

Mike Thelwall et al. (m.thelwall@wlv.ac.uk): Conceptualization (Equal), Data curation (Equal), Methodology (Equal), Writing—original draft (Equal); Emma Stuart (emma.stuart@wlv.ac.uk): Methodology (Equal), Writing—review & editing (Equal); Amalia Mas Bleda (Amalia.Mas@wlv.ac.uk): Methodology (Equal), Writing—review & editing (Equal); Meiko Makita (MeikoMakita@wlv.ac.uk): Methodology (Equal), Writing—review & editing (Equal); Mahshid Abdoli (m.abdoli@wlv.ac.uk): Methodology (Equal), Writing—review & editing (Equal).

References


I’m Nervous about Sharing This Secret with You: Youtube Influencers Generate Strong Parasocial Interactions by Discussing Personal Issues


This is an open access article under the CC-BY license (https://creativecommons.org/licenses/by/4.0/).