

The representation of gendered social actors across five manosphere communities on Reddit

Krendel, A., McGlashan, M., & Koller, V. (forthcoming 2022). The representation of gendered social actors across five manosphere communities on Reddit. *Corpora*, 17(2).

Abstract

This paper investigates the representation of gendered social actors in a specialised corpus of 10.9 million words collected from five Reddit communities associated with the so-called manosphere: incels (involuntary celibates), Men Going Their Own Way (male separatists), pick-up artists, men's rights activists, and a group dedicated to wider discussions of 'red pill' philosophy. 34 gendered social actor terms were identified as key-key-words across the manosphere corpora. Both male and female social actors are referenced using relational terms, while the latter are also referenced using derogatory terms and the former are referenced using terms for kinship and in-group identification.

We then analyse the consistent collocates (Baker et al., 2008) of the four most frequent gendered social actor terms (*women*, *girls*, *men* and *guys*), to establish the topics, descriptions, and actions associated with the social actors across the five groups. Gendered social actors were constructed in essentialist dichotomies, with women and girls, although objectified and passivated in dating/sexual contexts, being represented as violent towards male social actors and as holding a privileged position over men in wider society.

The anti-feminist ideology reflected in manosphere discourse can be seen as a more extreme version of mainstream discourse, into which it may be re-imported.

Keywords: Reddit, manosphere, key-key-words, consistent collocates, gender, social actor representation

1. Introduction

This paper investigates the representation of gendered social actors across a corpus representing the so-called 'manosphere'. The 'manosphere' refers to a broad online network of websites and platforms where users share anti-feminist content and essentialist views about gender, and participate in 'networked misogyny', organising campaigns against feminists (Marwick and Caplan, 2018). Five subcommunities have been identified consistently in the literature (Ging, 2017; Jane, 2018; Ribeiro et al., 2020): men's rights activists, Men Going Their Own Way (male separatists: henceforth MGTOW), pick-up artists (PUAs), involuntary celibates (incels), and individuals who participate in manosphere discussions on what is known as 'red pill' philosophy but without identifying with a specific subgroup. These communities each have corresponding individual subreddits (i.e. dedicated spaces for the discussion of manosphere topics on the content aggregation and discussion site Reddit), and

past literature has investigated the anti-feminist and sexist aspects of individual subreddits in qualitative ways (e.g. van Valkenburgh, 2018 on /r/TheRedPill; Lumsden, 2019 on /r/MensRights).

The manosphere has been widely recognised as a ‘toxic’ community, with links to anti-feminist harassment campaigns (Jane, 2018), wider right-wing political movements (Lewis, 2019), and offline violence. The incel community in particular has received much media coverage, as some incels have claimed that frustration at their own lack of romantic and/or sexual relationships had motivated them to commit murder (Rodger, 2014). The perpetrators who have received the most media attention are Elliot Rodger in California in 2014, who killed seven people and injured fourteen, and Alek Minassian in Toronto in 2018, who killed ten people and injured sixteen. Approximately a dozen other such attacks have occurred, with one occurring as recently as May 2020 in Toronto being the first incel attack to be characterised as domestic terrorism (Cecco, 2020). Furthermore, Adnan Ahmed, a self-styled PUA from Glasgow, was jailed for two years in 2019, as using PUA tactics on women and girls in public constitutes ‘threatening behaviour’ (BBC, 2019). Although only a small number of offline manosphere actions have been criminally prosecuted, these examples suggest that an underlying ideology about gender relations across the manosphere could contribute to these behaviours.

Although some recent studies in corpus linguistics have focussed on the manosphere (see Section 2), the majority of these studies have concentrated on a single community. While each of these communities draws on essentialist views about gender to legitimise different approaches to women (e.g. avoidance or serial seduction), by considering the manosphere as one community composed of similar but distinct parts, we interrogate Marwick and Caplan’s (2018:553) claim that groups within the manosphere are ‘brought together by a *common language* that orients them in opposition to the discourse and rhetoric of feminism’ (emphasis added). Thus, we aim to uncover its underlying ideology, here defined as ‘a (metaphorical) network of beliefs that gives rise to expectations, norms and values about events, ideas and people’ ([Author], 2014:239). Ideology can be seen as the ‘common ground’ (Clark, 1996) that is shared by the participants. In particular, we investigate how a *common language* is used to represent, describe, and evaluate male and female social actors. As such, the analysis aims to uncover what ideologies around these gendered social actors are prevalent in the manosphere.

[Author] (2014:150) argues that ‘repeatedly exposing text recipients to certain [socio-cognitive representations] transported in texts, under similar conditions of reception, may help to align recipients’ cognition with that of the text producer’. This has also been described by corpus linguists such as Baker (2006:13) as ‘the incremental effect of discourse’, and has been examined using techniques such as keyword and collocation analysis. Doing so allows us to infer the ideology that underlies the representation of gendered social actors across the manosphere. To this end, we examine gendered social actor key-key-words (words which are key across a number of texts; Scott, 1997) shared across five manosphere subreddits along with their consistent collocates (words which collocate across subcorpora, hereafter referred to as c-collocates; Baker et al., 2008). To structure our findings, we employ van Leeuwen’s (2008) social actor framework to categorise these key-key-words, and then inductively categorise their c-collocates into topics, descriptions, and actions. We have organised this investigation around two research questions (RQs):

RQ1) How are gendered social actors represented across five manosphere communities on Reddit?

RQ1a) What words are used to identify gendered social actors?

RQ1b) When gendered social actors are identified, what topics, descriptions, and actions are they associated with?

RQ2) What ideologies of gender are social actor representations in the manosphere related to?

The remainder of the paper is organised as follows. Section 2 provides an overview of previous research which investigates representations of male and female social actors using corpus-based approaches, as well as the growing body of research which applies corpus linguistic methods to manosphere datasets. Sections 3 and 4 outline our Reddit data collection process and the procedure of calculating key-key-words and c-collocates, along with the categorisation systems used for analysis. We first present our results in Section 4 and further discuss our findings in Section 5. In Section 6, we conclude by explicitly answering our RQs and looking at the implications of our study.

2. Corpus-based research on gender representations

Collocation analysis has been a central method used in the past literature to investigate gendered social actors in general corpora of English. This is, for instance, evidenced by Romaine's (2000:103) analysis of the adjectival collocates of *bachelor* and *spinster* in the British National Corpus, in which she argues that the collocates of *spinster* are more likely to be negative, e.g. *jealous*. Ultimately, the collocates indicate that there are negative discourse prosodies around nominal terms for women and that discourse prosodies are typically more positive for nominal terms for men. Similarly, Pearce's (2008:8, 12) analysis of the pre-modifying adjectival collocates and verbal collocates of *man* and *woman* in the British National Corpus using SketchEngine reveals that *man* is more likely to collocate with terms denoting physical size (and implied strength), such as *broad-shouldered*, while *woman* is more likely to occur with terms denoting social categories, such as *married*. Pearce (2008) also notes that *man* is more likely to collocate with terms denoting power and success.

Elsewhere, Sigley and Holmes' (2002:151) analyse the collocates of *boy* and *girl* in five general corpora of British, American, and New Zealand English, and found that collocates of *girl* typically relate to their physical attractiveness and domestic skills, whereas the collocates for *boy* relate to a range of age, appearance, and behaviour descriptors. Also looking at the terms *boy* and *girl*, Baker (2014) explores the collocates of these gendered nouns in the ukWaC corpus of British websites. His analysis examines different verb collocates of these terms, and in particular focuses on whether these gendered terms are the agent or patient of the verb collocates. He finds that *girl* is more likely to collocate with terms such as *rape* and *abducted* in the patient position (i.e. they were more likely to be represented as victims), while *boy* is more likely to occur as agent with collocates relating to physical violence (such as *beat*).

Furthermore, utilising corpus methods in tandem with other discourse analytical frameworks, Caldas-Coulthard and Moon (2010) and Moon (2014) both analyse the adjectival collocates of *man/men*, *woman/women*, *girl* and *boy* in the 450-million-word Bank of English corpus. Both utilise van Leeuwen's (2008) social actor representation framework to structure their results, and find that female social actors are typically described in terms of their sexuality

and physical appearance, whereas male social actors are described in terms of strength, physical activity, and ability.

While the majority of the past literature foregrounds differences between male and female social actors, Taylor (2013) instead analyses the similarities in the c-collocates of *boy* and *girl* in three UK broadsheet newspapers, using the SiBol 93, SiBol 05, and Port 2010 corpora. She notes that about a third of a total 119 c-collocates were shared between *boy* and *girl*, and that sexual relationships are prevalent for both *boy* and *girl*, although more frequently indicated alongside *girl*.

Broadly, all of the comparative studies discussed here conclude that male social actors are represented as more agentive and occupying more powerful positions than female social actors. Contrastingly, female social actors are represented in terms of physical appearance and their relationships to male social actors. This demonstrates that binary distinctions between male and female social actors are constructed in general language use, which could incrementally affect the way gendered social actors are conceptualised in the minds of other speakers, with such concepts in turn influencing language use.

2.1 Corpus linguistic research into the manosphere

Corpus methods are increasingly being applied to studies of the manosphere, with most studies concentrating on a single community or facet of the manosphere and rarely exploring multiple communities. Studies of PUAs include Dayter and Rüdiger's (2016) work on 37 'field report' postings from PUA forums as well as Wright's (2020) analysis of the frequency and collocates of lexis denoting resistance in a 26-million-word corpus of PUA discussion forum data. These studies have found that pseudo-technical language is used to bring credibility to PUA techniques and create emotional distance between PUAs and the women they target, and that in-group experiences with women are framed as sequences of complicating actions to be overcome. Additionally, Lawson and [Author]'s (2017) keyword and key-keyword (Scott, 1997) analysis of posts collected from three PUA/seduction subreddits and r/TheRedPill (which consists of manosphere members unaffiliated with a specific sub-group) reveals that gendered social actors, swear words and taboo terms, and affective/mental processes are referenced across the four subreddits, with female social actors referred to in derogatory and sexualised ways (e.g. *bitch* and *hb*, denoting 'hot babe').

Focussing on an incel subreddit, Heritage and [Author] (2020) collect a corpus of 50 threads, comprising approximately 67,000 words of running text from both original posts and comments. They analyse which social actors terms are key compared to the American English 2006 corpus (Potts and Baker, 2012), the frequencies of these terms, and how social actors are appraised, using Martin and White's (2005) appraisal framework. Much like Lawson and [Author] (2017), Heritage and [Author] (2020) note the use of pejorative terms to refer to female social actors, although male social actors are also referred to in this way in the incel dataset, as they are placed in a hierarchy which includes terms such as *manlets*, i.e. men who are judged as less capable than men who enact ideals of hegemonic masculinity (see Connell and Messerschmidt, 2005). In terms of appraisal, Heritage and [Author] (2020) show that male social actors are judged as incapacitated and unhappy, whereas female social actors are perceived as having the capacity to hurt men, as well as being dishonest and immoral.

[Author] (2020) uses the same appraisal framework in her analysis of the gendered social actor keywords *men*, *women*, *guys* and *girls* (compared to a 1.65 million-word sample of the Corpus of Contemporary American English; Davies, 2010) in a corpus of 214,269 words made up of posts and comments from the *TRP* subreddit. This analysis is supplemented by an analysis of pre-modifying adjectival collocates and verbal collocates, using the SketchEngine Word Sketch tool (Kilgarriff et al., 2014), to check that the qualitative findings obtained are representative of the whole corpus. She finds that female social actors are dehumanised and sexually objectified, are represented as wanting hostile behaviour from male social actors, as well as being dishonest and immoral. On the other hand, male social actors are represented as unhappy and insecure. [Author]'s (2020) findings therefore mirror Heritage and [Author]'s (2020) results for incel discourse.

In sum, various studies have explored gender in the manosphere with corpus linguistic tools and noted that female social actors are referred to in derogatory ways in multiple subgroups. Furthermore, Marwick and Caplan (2018: 553) claim that the manosphere shares a common language, a claim which is echoed by Bates (2020) albeit not from a linguistic perspective. However, there remains a gap in the research in that we currently do not have statistical evidence of the language, beliefs and attitudes which unite all five subgroups, a research gap addressed in the present study.

3. Data Selection and Collection

Although there are dedicated manosphere websites, the popularity of such websites has not been established. By contrast, past literature on the manosphere (e.g. Lawson and [Author], 2017; Lumsden, 2019) has revealed that Reddit is used by multiple manosphere communities which attract large subscriber numbers. Reddit is a content aggregation and discussion website, where users can create and subscribe to dedicated topic-specific communities, known as subreddits. In these subreddits, users post topic-specific content including text posts, pictures, and content from other websites. Users can also comment on each other's posts.

We selected five subreddits for our study which pertain to the five parts of the manosphere discussed in Section 1: *r/MensRights* corresponds to men's rights activists, *r/MGTOW* to Men Going Their Own Way, *r/sexduction* to PUAs, *r/braincels* to incels, and *r/TheRedPill* to manosphere members unaffiliated with a specific sub-group. At the time of writing, *r/MensRights* has 299,220 subscribers, and *r/sexduction* has 613,638 subscribers. As the remaining three subreddits are either quarantined or banned, their subscriber counts are hidden, but at the time of quarantine or ban, respectively, *r/TheRedPill* had approximately 300,000 subscribers, *r/MGTOW* had approximately 140,000 subscribers, and *r/braincels* had approximately 80,000 subscribers.

On Reddit, users can upvote and downvote posts and comments, to show that they like or dislike the post or comment respectively. If a post has many upvotes, this indicates that the post is popular, whereas the opposite is true for downvotes. We collected comments on the 200 most upvoted (and therefore most popular) submissions from the five subreddits since their inception. The original posts were not considered in this analysis, as many of them are not made up of original text, but use links to external websites to generate discussion in the comments section. The Reddit manosphere corpus was collected using the Python Reddit API Wrapper (PRAW, 2020). By collecting the top 200 comment threads in each subreddit (a total of 1,000 threads), as determined by the Reddit upvotes, we generated a corpus of 10.9

million words, which contains five subcorpora. By building subcorpora of the most popular posts in each subreddit, we attempt to capture the widely accepted ‘common ground’ within that specific subcommunity as well as across the manosphere more generally. Details about corpus size are presented in Table 1. When quotations are taken from these threads to illustrate our findings, the number assigned to the thread and the originating subreddit is given (e.g. 184_MGTOW).

Subreddit	Filename convention	Number of threads	Tokens
r/TheRedPill	TRP	200	2,977,113
r/MensRights	MR	200	5,019,556
r/MGTOW	MGTOW	200	888,930
r/seduction	SED	200	1,042,601
r/braincels	BRA	200	973,334
		1,000	10,901,534

Table 1: The Reddit manosphere corpus

As we used the Reddit API (application programming interface) to collect the data used for this study, we collected our data in accordance with the Reddit API Terms of Use. These Terms state that if one uses the Reddit API to collect user content from the site, Reddit grants one ‘a non-exclusive, non-transferable, non-sublicensable, and revocable license to copy and display the User Content using the Reddit API’ (Reddit API Terms of Use, 2016). Thus, even though the Reddit User Agreement (2021) states that users have ownership rights to their content, Reddit does not require API users to obtain consent from the Reddit users they collect data from. Furthermore, considering copyright, Reddit adheres to US copyright law, which only protects works which contain creative expression, and not works which solely constitute facts or ideas. As the Reddit posts and comments used in this study are not creative works, their use does not constitute copyright infringement according to the DMCA (Digital Millennium Copyright Act).

However, the researchers are aware of ethics guidance from the Association of Internet Researchers (franzke et al., 2020) which notes that obtaining informed consent from research participants is often best practice. Although some researchers seek informed consent from the online communities they wish to research (e.g. Mackenzie, 2017), this is not necessarily appropriate when researching communities which could pose a security risk to the researchers. Indeed, those who have researched the manosphere (e.g. Rüdiger and Dayter, 2017) acknowledge that making oneself known to such communities could result in the researchers being harassed and targeted online. Such incidents have led to the most recent ethical guidelines from the Association of Internet Researchers (franzke et al., 2020) explicitly recognising that in some instances, researcher safety must come at the expense of obtaining informed consent. Thus, we chose to not obtain informed consent from the moderators of the five manosphere communities we research.

4. Methodology

Our methodology comes in three parts: we (i) identified key-key-words shared across five manosphere subreddits (Table 1) and analysed them using van Leeuwen’s (2008) social actor framework to ascertain gendered social actors; (ii) identified c-collocates of these shared social actor key-key-words occurring at least five times in each of the five subcorpora, and; (iii) inductively categorised these c-collocates into topic indicators, descriptions of social

actors, actions either undertaken by the social actor (occurring to the right of the node) or actions in which the social actor is passivated (occurring to the left of the node), and miscellaneous c-collocates.

4.1 Identifying shared key-key-words across five manosphere communities

Keyword analysis aims to identify lexical items (types) used in a target corpus at an unusual relative token frequency when compared against another (usually larger) reference corpus. Egbert and Biber (2019: 88) suggest that keywords identified in this way are strong indicators of a target corpus' "content-distinctiveness". As such, keywords have been described as "lexical signposts" (Baker, 2004a: 90) providing a "rapid and useful way of directing researchers to elements in texts that are unusually frequent (or infrequent)" (Baker, 2004b:348). Although widely used to identify content-distinctive lexis, keyword analysis has been criticised for overplaying lexical differences and obscuring the potential lexical similarities between corpora (Baker, 2004b: 346) as well as for treating the texts within a target corpus as a single, homogeneous whole. In response, methods for keyword analysis have been elaborated to identify content-generalisable keywords, i.e. keywords that are dispersed across a target corpus' composite texts. Notable methods include Egbert and Biber's (2019) text dispersion keyness and Mike Scott's (1997) key-key-word approach; we have adopted the latter in this paper.

To identify key-key-words, we first produced a list of positive keywords - lexical items used at a significantly higher frequency than is found in a reference corpus - for each of the 200 threads that make up our five separate corpora described in Table 1 (1000 total). For our reference corpus, we used the WebCorp Mini-Web corpus (2010) consisting of 100,000 randomly sampled English language webpages (totalling 339,907,995 tokens) collected between 2000 and 2010. This was due to its size and composition, in that it draws on data written for online consumption, making it register appropriate, and across a range of different topic domains. These keyword lists were compiled using a script written in R following the guide set out on the UCREL log-likelihood and effect size calculator webpage (Rayson, 2008). All keywords identified in these 1000 keyword lists were measured for statistical significance using the Log-Likelihood (LL) statistic and were only deemed to be key if they met a minimum threshold of $LL \geq 15.13$ ($p < 0.0001$). Following the production of keyword lists, we then focussed our attention on the production of key-key-word lists for each of our five manosphere subcorpora. Key-key-words were identified by following Scott's (1997) approach which notionally tallies the number of times a keyword is found to be key across each individual text in a corpus. As Scott (1996) documents in the original manual for WordSmith Tools, 'a "key key-word" is one which is "key" in more than one of a number of related texts. The more texts it is "key" in, the more "key key" it is'.

In our study, we produced a list of key-key-words for each subcorpus (a tally of the keywords found in each of the 200 threads) and ranked keywords by their key-keyness (i.e. *dispersion*, or how many times they occurred as keywords across the 200 threads). We then selected only those 528 key-key-words that were present in all of the five key-key-word lists. We argue that these 528 shared key-key-words are both distinctive of the respective subreddits they originate from (as they are significantly frequent across the subreddits) as well as

generalisable to the language across these five manosphere communities. As such, they provide a basis for interrogating Marwick and Caplan’s (2018: 553) claim that the manosphere is unified through the use of a “common language”.

To address our specific focus on gendered social actor representation across Reddit manosphere communities, we then identified the nominal items from the key-key-words list which indexed gendered social actors either explicitly (e.g. *men*) or implicitly (e.g. *whore*). Where the gender of a social actor was indexed implicitly, gender was determined by checking the concordance lines for patterns in the co-text (e.g. “women are whores” (184_MGTOW)), and by applying contextual knowledge of the manosphere (e.g. Ging, 2017) to determine which gender is indexed (e.g. *incel* as male).

We then sorted the gendered key-key-words using van Leeuwen’s (2008:42) social actor analysis framework. In doing so, we focussed on categorisation, which considers how social actors are represented in terms of the qualities they share with others. Categorisation is further split into appraisal (evaluation), functionalisation (what a social actor does) and three subtypes of identification (what a social actor more or less permanently is):

- Classification - social actors are identified through demographic information, e.g. gender, age, ethnicity
- Relational identification - social actors are identified through their relationships with others
- Physical identification - social actors are identified through physical characteristics

Of the 528 key-key-words shared by the subcorpora, 34 refer to nominal gendered social actors, and are displayed in Tables 3 and 4. Cases where *bitch* was a verb and where *male* and *female* modified a head noun (e.g. ‘female friend’) were filtered out using a part-of-speech tagger (TreeTagger, which is included in #LancsBox; Brezina et al., 2018), so that only nominal cases of *bitch*, *male* and *female*, and the collocates and c-collocates which met our statistical threshold, were considered. Across the corpus, male social actor terms were used 94,605 times in a mean of 458.5 comment threads, and female social actor terms were used 98,953 times in a mean of 429.9 comment threads. Thus, discussions about female social actors were slightly more prevalent than discussions of male ones.

Key-key-word	Frequenc y	Dispersion	Number of collocates	Number of c- collocates
<i>men</i>	43155	892	1159	104
<i>man</i>	17812	941	549	5
<i>guy</i>	11161	898	387	96
<i>guys</i>	9535	885	348	97
<i>incel</i>	2035	311	99	0
<i>male</i>	1849	491	73	10
<i>dude</i>	3799	783	176	31
<i>dudes</i>	1061	449	67	9
<i>bro</i>	1711	598	81	16
<i>bruh</i>	138	110	0	0
<i>boyfriend</i>	756	287	37	8
<i>bf</i>	277	134	10	3
<i>cuck</i>	686	281	34	0

<i>pua</i>	468	175	22	0
<i>neckbeards</i>	135	85	0	0
<i>niceguys</i>	27	16	0	0

Table 3: Key-key male social actor terms, their frequency and dispersion (how many texts they are key in) across the corpus, and their (c-)collocates, ordered by frequency of and within terms

Key-key-word	Frequency	Dispersion	Number of collocates	Number of c-collocates
<i>women</i>	55197	947	1260	145
<i>woman</i>	14768	858	423	78
<i>girl</i>	9049	826	328	84
<i>girls</i>	8536	800	321	86
<i>girl's</i>	220	132	8	1
<i>bitch</i>	2026	564	89	19
<i>bitches</i>	875	380	61	6
<i>girlfriend</i>	1171	435	58	15
<i>gf</i>	561	273	27	7
<i>girlfriends</i>	311	197	20	4
<i>females</i>	1570	463	65	11
<i>female</i>	1095	399	49	4
<i>chick</i>	964	380	54	8
<i>chicks</i>	711	325	44	2
<i>slut</i>	648	246	33	4
<i>whore</i>	598	250	32	2
<i>whores</i>	416	204	18	0
<i>lesbians</i>	237	59	11	0

Table 4: Key-key female social actor terms, their frequency and dispersion (how many texts they are key in) across the corpus, and their (c-)collocates, ordered by frequency of and within terms

Male and female social actors are categorised in different ways in the key-key-words, as shown in Table 5.

			Male	Female
Categorisation	Identification	Classification	<i>men, man, guy, guys, male, dude, dudes, incel, pua</i> (9)	<i>women, woman, girl, girls, girl's, females, female, chick, chicks, lesbians</i> (10)
		Relational identification	<i>bro, bruh, boyfriend, bf</i> (4)	<i>girlfriend, gf, girlfriends</i> (3)
		Physical identification	<i>neckbeards</i> (1)	
	Appraisal		<i>cuck, niceguys</i> (2)	<i>bitch, bitches, slut, whore, whores</i> (5)

Table 5: Representation of gendered social actors through key-key-words (see van Leeuwen, 2008: 52)

Male and female social actors are most commonly classified in terms of their gender (e.g. *men*, *women*), as well as by their group membership within the manosphere (*incel*, *pua* [pick-up artist]) and, for women, with regard to their sexual identity (*lesbians*). The second most frequent categorisation for female social actors is by way of negative, sexualised appraisal or evaluation (*bitch(es)*, *whore(s)*, *slut*). This is in stark contrast to only two negative evaluative categorisations for male social actors (*cuck* [cuckold], *niceguys*), only one of which is sexual in nature (see Marwick and Caplan, 2018). This finding corroborates previous studies of both general and manosphere corpora that show women and girls to be represented in negative and sexualised terms more often than men (e.g. Romaine, 2000; [Author], 2020). Gendered social actors are also relationally identified as *boyfriend/bf* and *girlfriend(s)/gf*, but key-key-words which refer to metaphorical kinship (*bro* and *bruh*) are specific to male social actors, as is the one, derogatory instance of physical identification (*neckbeards*). While this overview demonstrates the negative bias of manosphere discourse, it also shows important differences in the representation of male and female social actors.

4.2 Identifying c-collocates

Following the identification of gendered social actors, c-collocates (Baker et al., 2008) of the respective key-key-words were calculated by first using the GraphColl tool in #Lancsbox (Brezina et al., 2018) to determine the collocates of each relevant social actor in each of the five subcorpora, and then using Microsoft Excel to identify which of these collocates were consistent across subcorpora. Following the guidelines set out in Gabrielatos and Baker (2008), the minimum frequency that a collocation needed to occur in each corpus was five. Thus, a collocation needed to occur at least five times in each of the five subcorpora (and therefore a minimum of 25 times) to be considered a c-collocate. Calculating c-collocates filters out ‘seasonal collocates’ (Baker et al., 2008), which are collocates specific to one subcorpus and, therefore, unrepresentative of typical language use in the corpus as a whole. This helped to ensure that the larger subcorpora (*Men’s Rights* and *The Red Pill*) were not over-represented. By conducting such an analysis, we aim to reveal ‘the associations and connotations [words] have, and therefore the assumptions which they embody’ (Stubbs, 1996:172).

C-collocates were identified as those words occurring within a span of five words from the node word (5L/5R), with the collocation occurring at least five times in each of the five subcorpora, and with an MI score of at least 3 and a T-score of at least 7.5. These statistical thresholds are in line with, or considerably higher than, the accepted field standards of an MI score of 3 and a T-Score of 2, which Hunston (2002) recommends. The cut-off points we use were informed by Durrant and Doherty’s (2010) lexical decision task study, in which they found that collocate pairs with MI scores of at least 6 and T-scores of at least 7.5 could be considered psychologically real. However, we found that an MI threshold of 6 did not yield enough c-collocates to allow for an extended discussion and therefore chose to keep a minimum MI score of 3, while acknowledging that the c-collocates with the highest MI scores are expected to best reflect psychological reality, although this is impossible to determine without experimental data. This method resulted in 379 c-collocates for the key-key-words denoting male social actors, and 476 c-collocates for the key-key-words denoting female social actors.

For reasons of space, we investigate the c-collocates for the four most frequent gendered social actors: *women*, *girls*, *men* and *guys*. In total, these social actors have 432 c-collocates. Only the plural forms are selected for analysis, to investigate how collective gendered identities are represented, as opposed to individualised ones.

4.3 Categorising c-collocates

Thirdly, we placed c-collocates into five categories: topic indicators, descriptions, actions to the left of the node (reflecting social actor as a patient), actions to the right of the node (reflecting social actor as an agent), and a bin category (cf. Table 5). Where a c-collocate can be placed in more than one of the categories described in Table 5, it is categorised multiple times. The action c-collocates were sorted using the GraphColl tool in #LancsBox (Brezina et al., 2018), which notes whether collocates occur more frequently to the left or right of the node. To interpret how these c-collocates were employed in the dataset, we used WordSmith 7 (Scott, 2016) to generate 25 randomised concordance lines for each c-collocate (using the ‘Reduce to N’ function), with a window of 150 characters either side of the node for additional co-text. There are two reasons why we chose 25 random concordance lines as the number of lines to closely read for each c-collocate for two reasons. Firstly, 25 is the minimum frequency of the c-collocate together with the node word across the corpus, and so 25 was the minimum number of concordance lines which could capture how the c-collocate was used across all five subcorpora. Secondly, given that the four social actors of interest had a total of 432 c-collocates, this required the manual reading of 10,800 concordance lines (25 concordance lines per c-collocate) between three researchers. This was deemed to be a maximum manageable workload for the researchers. In the following section, we discuss the results generated from the application of these methods.

C-collocate category	Definition
Topic indicators	nominal and adverbial items which denote a particular topic being associated with the node
Descriptions	adjectival and adverbial items which quantify and describe the node
Actions to the left of the node	verbs which occur to the left of the node
Actions to the right of the node	verbs which occur to the right of the node
Miscellaneous	Grammatical particles; auxiliary and modal verbs; items which indicate argumentation, stance, or intensification

Table 6: C-collocate categorisation system

5. Results

Figures 1-4 display the 432 categorised c-collocates of *women*, *girls*, *men* and *guys* in different shades, with their position in relation to the node and their MI score given in the bottom left and right-hand corners of each square respectively. Squares are scaled and ordered by MI scores; squares with larger MI scores are positioned to the top left of a category and are larger, squares with smaller MI scores are positioned to the bottom right of a category and are smaller. In the following discussion, we will distinguish between key-key-words and c-collocates by using italics to indicate key-key-words whereas double quotation marks will be used to indicate c-collocates. Furthermore, as it is not feasible to discuss each of the 432 c-collocates in detail in this paper, we discuss those c-collocates which constitute evidence of a consistent theme across multiple c-collocates. This is to ensure that we discuss the themes which are referenced most consistently using the c-collocates. We also consider the c-collocates which are shared between social actors, to serve as points of comparison between the social actors, as well as c-collocates which constitute comparisons and binary opposites (e.g. “hard” for *men* versus “easier” for *women*). Furthermore, when considering the social actors in isolation, the c-collocates with the highest MI scores are given the most space in the discussion.

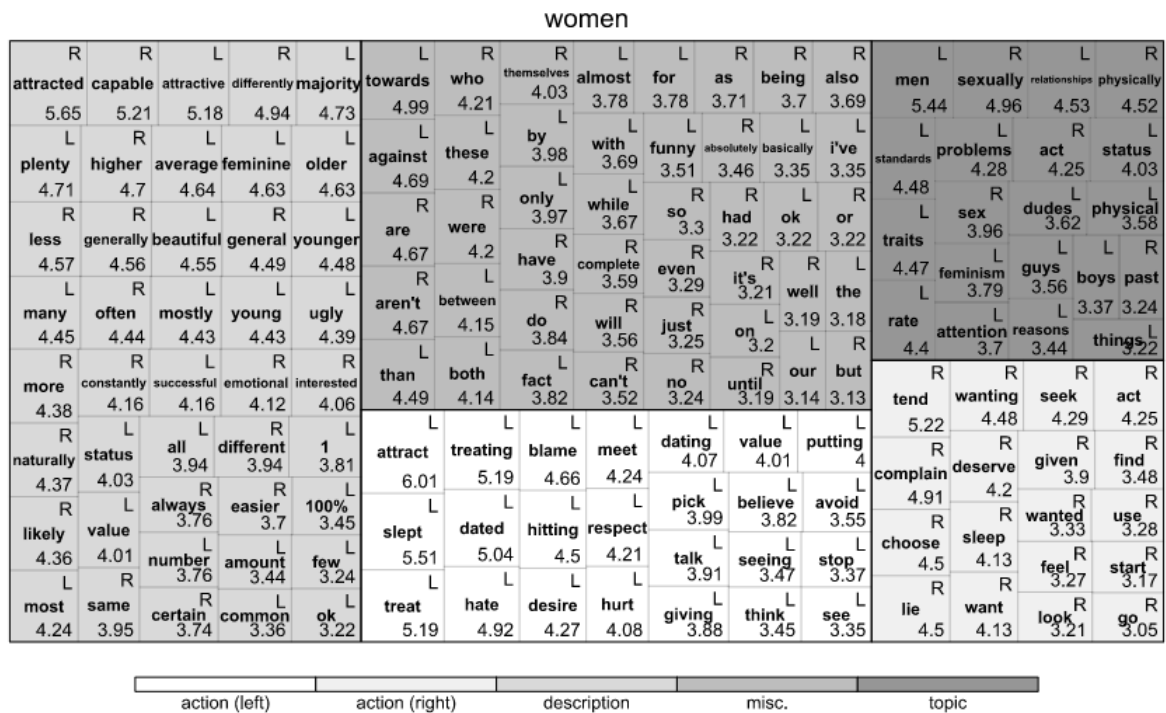


Figure 1: 145 c-collocates of *women*

girls

L	L	R	R	L	L	L	L	L	L	R	R		
i've	these	who	around	those	with	hot	young	fat	high	old	hit	love	
4.75	4.74	4.46	4.43	4.33	4.33	7.13	6.1	5.82	4.73	4.66	5.33	4.45	
R	R	L	R	L	L	L	L	L	R	R	R		
aren't	don't	only	too	had	from	really	most	pretty	lot	two	always	start	want
4.32	3.88	3.87	3.86	3.77	3.73	3.71	4.53	4.41	4.3	4.28	4.17	4.4	4.2
R	R	L	R	R	R	R	L	L	L	L	R	R	
are	over	when	they're	do	also	and	many	other	few	different	all	go	like
4.27	3.69	3.63	3.62	3.61	3.6	3.59	4.52	4.38	4.15	4.05	3.81	4.06	4.06
L	L	R	L	R	R	R	L	L	L	R	R	M	
up	than	at	how	on	they	still	white	year	some	same	never	give	think
4.27	3.69	3.59	3.55	3.55	3.55	3.53	4.46	4.36	4.13	3.75	3.31	3.85	3.52
R	R	R	L	R	R	L	L	L	L	L	L	R	R
will	yeah	them	to	but	not	of	talking	get	know	girls	school	sex	
4.15	3.69	3.58	3.52	3.43	3.36	3.33	4.95	4.34	3.98	5.87	5.35	3.75	
R	L	L	R	L	R	R	L	L	M	L	L	L	
like	there	where	even	for	that	in	fuck	try	think	guys	thing	3.32	
4.06	3.67	3.58	3.51	3.31	3.25	3.22	4.54	4.2	3.52	5.76			
R	R	R	R	R	L	L	L	L	L	L	L	L	
were	after	just	because	their	the								
4.02	3.65	3.56	3.48	3.3	3.15								

action (left)
action (right)
description
misc.
topic

Figure 2: 86 c-collocates of *girls*

men

L	L	R	R	R	L	L	L	L	R	R	R	R
among	by	ones	themselves	their	as	for	gay	majority	differently	higher	boys	women
5.18	4.23	4.16	4.15	4.08	3.96	3.94	5.34	5.11	5.07	5.07	5.49	5.44
R	L	R	R	R	R	R	L	L	R	L	L	L
are	these	especially	also	should	being	they're	attracted	average	often	more	society	men
4.79	3.94	3.81	3.8	3.76	3.72	3.7	4.95	4.86	4.58	4.56	4.66	4.34
L	R	L	L	L	L	L	L	L	L	L	R	L
than	have	that	too	of	can't	yes	less	most	attractive	successful	relationships	problem
4.72	3.91	3.67	3.63	3.58	3.54	3.52	4.51	4.32	4.31	4.05	3.93	3.78
R	L	L	L	R	R	R	L	R	R	R	R	L
aren't	fact	about	there	around	be	in	many	different	number	better	sex	sexual
4.52	3.87	3.66	3.47	3.43	3.43	3.43	4.5	4.02	3.88	3.66	3.71	3.7
L	L	R	R	L	L	L	R	L	R	L	L	L
between	while	and	having	like	us	it's	naturally	other	usually	few	world	attention
4.47	3.86	3.64	3.44	3.36	3.31	3.28	4.35	3.98	3.61	3.24	3.56	3.35
L	R	R	R	R	R	R	L	L	L	L	R	R
both	do	don't	not	into	can	would	lots	lot	important	together	hard	
4.47	3.85	3.63	3.44	3.31	3.24	3.17	4.35	3.88	3.46	3.37	3.37	3.28
R	L	R	L	L	R	L	L	L	L	R	R	R
who	because	they	when	so	them	very	blame	hate	dating	act	getting	feel
4.39	3.84	3.63	3.44	3.31	3.23	3.1	4.96	4.45	3.33	4	3.42	
								think	see	want	get	
								3.4	3.26	3.63	3.39	3.22

action (left)
action (right)
description
misc.
topic

Figure 3: 104 c-collocates of *men*

guys

these	hey	who	those	aren't	here	tall	nice	short	plenty	talk	get	talking	
5.49	5.31	5.13	4.93	4.63	4.58	6.73	6.03	5.64	5.55	4.41	4.17	4.08	
are	i've	lol	doing	over	really	with	lot	many	young	other	look	getting	come
4.55	3.94	3.93	3.92	3.91	3.91	3.88	5.26	5.01	5	4.71	4.02	3.94	3.91
fuck	do	how	so	yeah	up	just	looking	some	good	all	trying	want	need
4.18	3.85	3.7	3.65	3.65	3.62	3.61	5.23	4.5	4.09	3.86	3.91	3.82	3.74
were	there	why	for	actually	than	of	most	two	better	same	saying	say	
4.07	3.76	3.61	3.55	3.54	3.53	3.48	5.14	4.38	3.81	3.69	3.6	3.57	
like	pretty	even	because	much	only	this	bad	different	more	find	got		
4.06	3.74	3.59	3.47	3.44	3.44	3.43	4.82	4.32	3.8	3.58	3.47		
too	have	their	mean	had	but	by	fucked	know	love	help	girls	sub	
4	3.73	3.59	3.47	3.42	3.33	3.32	4.82	4.24	4.12	3.94	5.76	4	
don't	out	about	on	them	that	the	seen	date	think	tell	guys	problem	
3.96	3.72	3.57	3.45	3.41	3.31	3.24	4.45	4.15	4.05	3.93	5.31	3.63	
				from	would	to				see		women	
				3.36	3.3	3.17				3.85		3.57	

action (left)	action (right)	description	misc.	topic
---------------	----------------	-------------	-------	-------

Figure 4: 97 c-collocates of *guys*

5.1 Essentialist dichotomies

Common to all social actors studied here is a strong association with topic c-collocates that reference other gendered social actors: the strongest topic c-collocate of both *girls* and *guys* is “girls”, for *men* the strongest c-collocate is “boys”, and for *women* it is “men”. The consistency of collocation between the terms under study with other gendered social actor terms suggests that within the manosphere, these gendered social actors are understood in terms of their (construed) relationships with each other and with other gendered social actors.

5.1.1 Representations of homogeneous groups

Relationships between gendered social actors are framed in terms of their differences, especially allegedly immutable differences between men and women. Such discursive constructions have been evidenced for more mainstream discourses as well, which have since been critiqued by feminist critical discourse analysts such as Mills (2008) for their essentialism and for perpetuating often harmful gender stereotypes. In our data, these beliefs are indexed by the prevalence of lexemes of DIFFERENT as description c-collocates for *girls*, *guys*, *men*, and *women*. For *women*, the description c-collocates “different” and “differently”, when found alongside the c-collocate “men” within a 5L/5R window, explicitly evaluates women and men as being biologically and socially different, e.g. “men and women deal with stress differently mentally and chemically” [121_TRP]. These differences can also be assumed as biological differences through the c-collocate “naturally” for both *men* and *women*, which is used to frame these differences as biologically determined and thus objectively innate (e.g. ‘women naturally have much higher standards than men’ [138_BRA]).

However, although infrequent, some comments reveal possible ideological contestation in the manosphere concerning biological differences. For instance, the comment ‘men and women are biologically different that is not sexist. Choosing to interpret that difference in such a way that woman don’t have the potential to do a job simply for being a woman is wrong’ [174_MR] shows a user who is explicit in their understanding of men and women being biologically different whilst also rejecting essentialist and deterministic interpretations of these differences.

Although the presence of “same” – an antonym of “different” – as a c-collocate for *girls*, *guys*, and *women* potentially suggests a focus on commonalities between gendered social actors (e.g. ‘we have the same problems as women’ [107_MR]), “same” is more often used to argue against the presence of such commonalities between women and men, commonly through negation (e.g. ‘I don’t think men and women are the same’ [114_BRA]). Moreover, notions of commonality and equality between men and women may come into conflict with essentialising discourses of gender, as in ‘in an ideal world men and women would be treated the same, but an ideal world excludes the things that exist in the real world. Men are larger, stronger, and more aggressive than women are’ [3_MR]. The user goes on to suggest that benevolent sexism is important for women who are potentially at risk from men ‘in the real world’.

Identities are further homogenised and distinguished using numerals (“1”, “one”, “two”), quantifiers (e.g. “lot”, “few”), comparatives (e.g. “less”) and superlatives (e.g. “most”), all of which are descriptive c-collocates of *girls*, *guys*, *men* and *women*. These forms are largely used to homogenise groups of social actors, with increasing intensity as they move from quantifier to superlative. Quantifiers such as “few”, which occurs only as a c-collocate of *men* and *women* (e.g. ‘I have however met quite a few men who’ve slept with many women’ [61_MGTOW]), and numeral c-collocates for *girls* and *guys* serve to aggregate individuals (van Leeuwen 2008: 38). Numerals, for example, do this through determination (‘Now I’ve moved to the big city ... I’ve banged two girls’ [87_TRP]).

Other c-collocates of the social actor key-key-words studied here are the quantifier “many” and its related gradable forms, i.e. the comparative “more” and the superlative “most”. Their presence potentially suggests a common lexico-grammatical choice for representing social actors across the manosphere as homogeneous. When acting as an adjective quantifying *men* and *women*, “many” can be used to distinguish men and women, particularly through (re-)establishing deterministic gendered traits and subsequent tensions. For example, one post represents men as being biologically determined to not raise children, as in ‘the entire genesis of child support etc was to avoid the biological imperative - men impregnating women, bouncing, not sticking around to help raise the child. The reason for the law is to protect women from what many men naturally want to do’ [110_BRA]. Furthermore, the superlative form “most” is associated with more severely reductive identity (re)presentations than the other forms and again relies on negative gender stereotypes of men and women (e.g. ‘I believe that while most women are better than most men at manipulation’ [80_TRP]).

However, there is evidence of a counter-discourse within the manosphere which challenges essentialist representations of women, such as in ‘it seems like most men on here hate women and dehumanize them like they’re a different species, but then I see posts where a lot of the same people are upset/depressed about not having a partner’ [9_BRA]. Not only does such a counter-discourse suggest the presence of multiple (and competing) discourses on gender in

the manosphere, but it also serves to acknowledge and confirm the presence of the derogatory and dehumanising discourses that we analyse in this paper.

In the following, we will look at how commenters in our data compare male and female social actors, and how relations between them are portrayed, before looking at the representation of such social actor groups on their own.

5.1.2 Comparisons of, and relations between, male and female social actors

The comparative forms “more”, “higher”, and “less” frame gendered social actors in ways that (re-)assert gendered conflicts by setting up gendered fields of action wherein actions carried out by and affecting men and women are constructed dichotomously (e.g. ‘women tend to earn higher grades and drop out less frequently than men’ [101_MR]). These constructions typically represent men as disadvantaged compared to women, as becomes evident through the analysis of c-collocates which discuss the treatment of male and female social actors in wider society.

Although *men*, *women* and *guys* are all conceptualised as having problems, as indicated by the topic c-collocate “problem(s)”, and although the action “treat(ing) *women*” refers to ‘treating women like shit’ [30_BRA] and ‘treating women like children’ [172_TRP], female social actors are at the same time represented as privileged over their male counterparts. This is visible via *women*’s descriptive c-collocates “ok” and “easier”, which express the way in which commenters see women being treated in wider society (e.g. ‘it’s way easier to be a woman in the United States in 2018’ [131_MR]). The verbs occurring to the left of *women* indicate that when patients, women are also the beneficiaries of undeserved privilege and support, as shown by “giving”, “putting”, and “respect” (e.g. ‘giving women awards because they are women’ [124_MR], ‘putting women on a pedestal’ [57_MGTOW], and ‘despite my lame attempts to love and respect women as delicate flowers, I learned the hard way they will eat your heart out’ [71_MGTOW]). This is also the case for verbs, such as “given” and “deserve”, which occur to the right of *women* (e.g. ‘women are given preferential treatment’ [74_MR] and the ascribed and rejected proposition that ‘women deserve things simply for being women’ [131_TRP]).

In contrast, the descriptor “capable” is used to refute the argument that women deserve special treatment (e.g. ‘women are perfectly capable of taking care of themselves’ [129_MR]). Furthermore, “complain” and “choose”, which occur to the right of *women*, are used to dismiss the grievances of women such as objectification (e.g. ‘everything women complain about is pure projection’ [40_BRA]), and under-representation in typically male-led industries (e.g. ‘women still don’t choose STEM careers’ [56_TRP]).

Contrastingly, in the *men* dataset, the struggles that men are seen as facing are presented as legitimate. For instance, “hard” is used to describe the perceived issues that men are confronted with (e.g. “men have an incredibly hard time getting support” [43_MR]) and the effort which men exert compared to women (e.g. ‘men have to work so hard to get laid’ [118_MGTOW]). This is also shown through the topic indicator “standards”, which refers to “double standards” in favour of women in 36 out of the 152 instances. In addition, the topic indicator “attention” and the descriptor “important”, c-collocates for both *men* and *women*, refer to women receiving more attention and being more important than men in various contexts (e.g. ‘Most intelligent guys will start to realise how bombarded most women are with male attention’ [102_SED]). Thus, issues for men are perceived as consistently downplayed in favour of those of women. Furthermore, the topic indicator “world” for *men* is

used to convey the perceived universality of such issues (e.g. ‘the plight of homeless men all over the world’ [123_MR]).

Men are represented as being subjugated by a coalition of women and those who are perceived as supporting the interests of women. Women and feminists passivate men, where “blame” and “hate” occur to the left of *men* (e.g. ‘most women today just try to shove the blame on men’ [107_TRP], and ‘extremist feminists hate men because that’s at the core of all feminist thought and ideology’ [104_MR]). “Society” as a whole is also framed as subjugating men, who are passivated in constructions such as ‘society is blatantly telling men not to maximise their value’ [28_TRP]. Conversely, when “hate” and “blame” occur to the left of *women*, commenters refute a presupposition that the in-group “hate” women and “blame” women for societal issues (e.g. ‘it’s not that we hate women, but we understand them now’ [39_TRP]).

As will be explored in more detail in Subsections 5.2 and 5.3 below, intimate relationships between male and female social actors are discussed via the topic indicators “sex”, “sexual”, “sexually” and “relationships”, which co-occur with *women*, *girls* and *men* in various combinations. As well as referencing consensual relationships, these topic indicators can be used to refer to abusive dynamics. For instance, “sexual(ly)” is used to reference harassment for both *men* and *women* (e.g. ‘the very real issue of sexual harassment of men’ [134_MR], and ‘women get sexually harassed on the street about their body’ [5_TRP]). Both male and female social actors are thus represented as victims as well as perpetrators, although instances where women are victims of abuse tend to be dismissed by the in-group. For instance, women are represented as lying about having experienced rape and sexual assault to harm male social actors, via “lie” occurring to the right of *women* (e.g. ‘all women lie about sexual assault and paternity when hit by hypergamous doubt’ [76_TRP]). Furthermore, although 26 out of 84 instances of “hit” to the right of *girls* are accounted for by the romantic/sexual phrasal verb “hit on”, “hit” more often references violence in the *girls* data (e.g. ‘girls can hit men too’ [45_MR]). In addition, women are also represented as perpetrators of abuse via the descriptor “capable” (e.g. ‘women are equally capable of violence as men’ [181_MR]). Thus, the in-group seek to correct a presupposition that women and girls are not as violent as male social actors. However, it should be noted that women are also represented as the victims of violence, as seen by “hitting” and “hurt” typically occurring to the left of *women* (e.g. ‘men can do evil things, they can hurt women beyond belief’ [189_MR]). However, such instances are typically downplayed in comparison to violence against men.

So far, we have established that, despite some evidence of a counter-discourse, male and female social actors are mostly represented as homogenous groups with immutable characteristics. Comparisons both express and reinforce this dichotomous concept of gender. The two groups are also portrayed as relating to each other, with women enjoying social privileges yet complaining and men being disadvantaged and facing unacknowledged struggles. While both women and men are represented as perpetrators and victims of violence, women are referred to as blaming, hating and framing men. We will now look more closely at how each of the two gendered groups is represented on its own.

5.2 Representation of *women* and *girls*

We established four main themes in the c-collocates of *women* and *girls*: sexual objectification, ascriptions of emotionality, discussions about ethnicity and age, and a less homogeneous representation for *girls* than for *women*.

5.2.1 Sexual objectification

Evidence for sexual objectification can be found in the proportion of c-collocates which account for physical descriptions of *women* and *girls* (29.7% and 25.6% respectively) compared to *men* and *guys*. This is reflected in descriptors such as “attractive”, “physically”, “beautiful” and “ugly” for *women*, and “hot”, “fat”, and “pretty” for *girls*. Objectification is also reflected in the actions which co-occur with *women* and *girls*, with 16.6% of *women* c-collocates reflecting the passivation of women, compared to 12.4% of c-collocates reflecting their agency. Actions to the left of *women*, such as “attract”, “slept”, “dating” and “hitting” (in the phrasal verb “hitting on”) represent them as passivated by male social actors in dating/sexual contexts. Similarly, “fuck” and “get” occur to the left of *girls*, with the latter used to reference acquiring a partner (e.g. ‘is that why I get no girls?’ [139_MR]). Users also share their personal approaches to, and advice for, dating girls using “try” and “talking” (e.g. ‘I would always try to get girls to go to multiple venues with me before getting them to mine’ [123_TRP] and ‘just practice talking to girls and it’ll become easy’ [200_SED]). In comparison, only one c-collocate, namely “sleep” for *women*, represents female social actors as active in these contexts, which indicates a lack of sexual agency for both women and girls. Although girls are more frequently agents than patients overall (9.3% and 7% of actions to the right and left respectively), it should be noted that girls are described more, and ascribed fewer actions, than women, and the term *girls* is frequently used to refer to either female social actors in their youth or in dating/sexual contexts. Contrastingly, the term *women* is used to discuss a variety of contexts, such as relationships, careers, and their treatment in wider society, as discussed in Section 5.1.2.

5.2.2 Emotionality

When women and girls are represented as agents, these actions often reference mental processes. For instance, when the affective and desiderative mental processes “love”, “want”, and “like”, as well as the cognitive “think”, occur to the right of *girls*, these actions are used to describe what girls allegedly seek in a dating/sexual context (e.g. ‘girls love to be approached, when it is done correctly and not forced’ [151_SED]). Similarly, desiderative and perceptive mental processes such as “want”, “seek”, and “feel” occur to the right of *women*, as does the descriptive c-collocate “interested”, which is used to ascribe attitudes to women (e.g. ‘women aren’t interested in the act of sex’ [103_BRA]). This is also done explicitly via the descriptor “emotional” for *women* (e.g. ‘women are emotional beings’ [35_SED]). Furthermore, the topic indicating c-collocates “standards” and “status” are used to discuss what women expect from men in a relational context. Women are represented as having “standards” which are ‘irrational’ [112_BRA], because they have ‘far higher standards than men’ [114_BRA], but also low, as in ‘women have zero standards’ [178_SED]. As for “status”, this refers to the type of men that women seek out for relationships, as in ‘women are attracted to high status men’ [121_SED], with “value” also occasionally used in this manner. Overall, this indicates that female social actors are often represented in terms of emotions and desires, although this is done more explicitly for women than girls.

5.2.3 Ethnicity and age

Furthermore, adjectives denoting ethnicity and age are unique to the description of women and girls. Although both women and girls are described in terms of age (via “young(er)” and “older” for *women*), youth is emphasised in the *girls* dataset via the c-collocates “high”, “school”, “year”, “old” and “young”. Indeed, 62 of the 92 instances of “old” occur in the trigram “year old girls”, which mainly refers to female social actors between the ages of 13 and 25. Moreover, 44 of the 112 occurrences of “school” are accounted for by the trigram “high school girls”. This trend is broadly in line with findings from general corpora of English (e.g. Sigley and Holmes, 2002), which note that *girls* is used to reference children, adolescents, and adults alike. In-group experiences with girls of this age are discussed (e.g. ‘the only girls I knew in school that had little to no hobbies were the ones that read *Cosmopolitan* in class’ [71_MGTOW]) and girls are represented both as victims of sexual abuse (e.g. ‘Larry Nasar abused young girls for decades’ [191_MR]) and as overtly sexual (e.g. ‘I’ve had 13 year old girls “flirt” with me before’ [7_MR]). However, it should be emphasised that these concordance lines do not encourage abusive or sexual behaviour towards young girls.

Furthermore, girls are the only social actor to be described in terms of ethnicity, specifically whiteness. White girls are ascribed a disparate range of personality traits, and described in terms of how attracted they are towards male social actors of other races, and vice versa, including girls desiring black and Indian men. The statistical significance of “white” suggests that the in-group acknowledge cultural differences in behaviour among girls, which was not found for women more broadly. However, it is unclear why other ethnicities are not similarly significant. On the other hand, the phrase ‘white girls’ could be used as a disparaging term, as it arguably holds this meaning in popular culture. Indeed, Slobe (2018) observes that ‘mock white girl’ performances in popular media portray white girls as excessively emotional, vapid, childish, cosmopolitan, and excessively consumerist. Thus, the prevalence of “white” as a c-collocate of *girls* in this dataset could indicate that the girls in question are perceived as having these personality traits, which are considered negative in both popular culture and the manosphere alike.

5.2.4 Heterogeneous representation of *girls*

Our c-collocate analysis reveals that *girls* are represented more heterogeneously than women. Unlike *women*, *girls* are quantified using small numbers (“one”, “two”), which indicates that users discuss individual experiences with girls (e.g. ‘I remember overhearing two girls talking at the bar’ [181_TRP]). Furthermore, although *girls* are described as immutably “different” from male social actors, “different” is more often used to discuss individual differences between girls than differences between girls and male social actors, as in ‘different types of girls and types of looks’ [117_SED]. Similarly, “same” is not used to denote homogeneity between girls as a wider demographic, but instead refers to similarities between girls and male social actors of similar attractiveness, as in ‘hot guys get away with being jerks the same way hot girls get away with being bitches’ [15_BRA]. Thus, as also found in [Author] (2020), *girls* are represented as individuals and in a somewhat less homogeneous manner than *women*, whereas *women* are referred to as an abstract, homogenous group.

5.3 Representation of *men* and *guys*

We established three main themes in the c-collocates for *men* and *guys*: discussions about dating and relationships (including what constitutes an attractive male social actor), ascriptions of personality and heterogeneity (especially in *guys*), and ways in which the in-group discuss interactions with each other and offer advice to other male social actors.

5.3.1 Dating and relationships

Fewer c-collocates reference dating and sexual relationships for *men* and *guys* than for *women* and *girls*, and do so in different ways for *men* and *guys*. The topic indicators “sex”, “sexual” and “relationships” co-occur with *men* but not *guys*, and only one action is ascribed to *men* regarding sexual relationships, namely “dating” to the left of the node, alongside the descriptors “attracted”, “attractive”, and “gay”. Although “gay” is used as neither a positive nor a negative term, its presence in this list suggests heteronormativity, as terms such as ‘straight’ or ‘heterosexual’ are not c-collocates. However, the descriptors of *guys* reveal in-group beliefs about what physical characteristics can lead to romantic and sexual success. The most statistically significant c-collocates for *guys* reference their physical appearance (“look”, “looking”), including height, where “tall” is conceptualised as positive and “short” as negative (e.g. ‘most women might prefer tall guys’ [176_BRA]). This suggests that physically imposing traits are associated with masculinity and sexual prowess. Furthermore, the actions “date”, “fucked”, and “love” typically occur to the left of *guys*, with female social actors as active in these constructions (although such agency in sexual contexts is not visible consistently in the form of c-collocates for either *women* or *girls*). Thus, as with the action c-collocates for *women* and *girls*, *guys* also lack sexual agency. However, it should be noted that *guys*, while described least among the four social actors (in 19.6% of c-collocates), are more frequently represented as agents than patients (13.4% vs 9.3% of c-collocates). The majority of these c-collocates frame *guys* as engaged in verbal activity (“talk(ing)”, “say(ing)”) and as benefitting (“get(ting)”, “got”), but without any overarching themes prevalent in the concordance lines. This suggests that despite lacking sexual agency, *guys* are nevertheless represented as more active than female social actors.

Furthermore, although the topic indicator “problem”, which co-occurs with *guys*, suggests a similarity in topic with *men* (i.e. perceived systemic issues), it in fact covers a wide variety of topics, such as dating strategies and types of guys (e.g. ‘the problem is guys think they have to compete with each other for these chicks’ [65_TRP]). Additionally, the c-collocate “different” is used to quantify guys (e.g. ‘they are usually fucking 3-5 different guys at once’ [71_MGT]), as opposed to referring to innate differences between guys and female social actors. Thus, the term *men* is used to discuss perceived systemic issues, whereas *guys* is used in a broader range of contexts, including relational ones.

5.3.2 Personality and heterogeneity

References to both positive (“successful”, “good”, “nice”) and negative (“bad”) qualities are made in the *men* and *guys* c-collocates, albeit more prevalently so for *guys*. “Good guys” and “bad guys” are presented as immutable identities (e.g. ‘you have your good guys and bad guys mixed up in life’ [24_MR]), and “nice” refers to out-group *guys* who act positively (but often disingenuously) towards female social actors to gain their approval (e.g. ‘manipulative guys that call themselves nice guys’ [37_MGTOW]). Additionally, the descriptor “successful” broadly refers to men achieving economic success and personal fulfilment, suggesting that despite perceived problems, men are nonetheless represented as able to achieve such success. Thus, commenters acknowledge a variety of personality types in male social actors, whereas this is not the case for all female social actors. It is moreover

noteworthy that “successful” as a c-collocate of *women* is often used to describe men’s sexual and romantic success with them: 48 of the 112 instances of “successful” are accounted for by the trigram “successful with women”. Unlike women, men are not referred to as the object of anyone’s sexual success.

Furthermore, like *girls*, *guys* co-occurs with quantifiers indicating smaller numbers (e.g. “some” and “two”), which indicates some level of heterogeneity in their representation, as individual experiences with guys are discussed. In comparison, more c-collocates describe *men* (23.1%), *men* are ascribed fewer actions, and *men* are equally represented as agents and patients in the action c-collocates (4.8% of c-collocates for actions to both the left and right of *men*). These findings, combined with the results discussed in Section 5.1, indicate that the term *guys* is used to discuss male social actors in terms of their qualities and actions, whereas the term *men* is used to discuss male social actors in terms of their perceived place in wider society.

5.3.3 In-group interactions and advice

Furthermore, a variety of interactions within the male in-group is discussed. The topic indicator “sub” for *guys* is used to discuss the characteristics of individual subreddits in a broadly negative manner (e.g. ‘too many guys in this sub have some sort of victim complex’ [99_MR]). Also, “come”, which occurs to the right of the node, refers to guys joining manosphere subreddits as a result of external problems (e.g. ‘a lot of guys come to [the seduction subreddit] because they have problems relating with women’ [153_SED]). Self-improvement and advice given in that respect are referenced and expressed via the action c-collocates “better” and “need”, which occur to the right of guys (e.g. ‘guys are trying to better their self worth’ [15_BRA]), and ‘you guys need some perspective’ [29_MR]). To the left of *guys*, “tell” and “help” serve a similar purpose (e.g. ‘Mhmm, I tell guys to become interested in many things.’ [183_SED]). Lastly, “love” is used to signal both sarcastic or genuine affection for fellow group members (e.g. ‘I love the way you guys always assume I’m a dude’ [91_MR]).

The descriptive c-collocate “better” also co-occurs with *men*, although in this context, it is used to signal a desire to improve conditions for men in general (e.g. ‘I want to help men be better and happier’ [104_MR]). Two contradictory strategies for achieving this are suggested via the c-collocate “together”. The first of these is fostering unity between men and women, as in ‘it’s so important for men and women to come together and talk human issues’ [172_MR], whereas the second is creating male-only spaces, as in ‘men need to stick together, and use our heads against these women taking advantage of us’ [149_MGT]. Contrastingly, the advice given in the *guys* concordance lines focus on self-improvement or refer to specific individuals, whereas the solutions in the *men* concordance lines pertain to gender relations more broadly.

To summarise subsections 5.2 and 5.3, we can say that both female and male social actors are represented in terms of their physical appearance, with an additional focus on sexual attractiveness for women and girls. The latter are also largely passivated in sexual and romantic contexts, while men are portrayed as successful in sexual, professional and economic terms, despite their perceived social disadvantage (see subsection 5.1.2). Finally, and despite perceiving gendered social actors as two homogenous groups when comparing them, commenters reference a variety of personalities for male in-group members and also allow for men to change.

In the final section of this paper, we will now answer our RQs and point out possible directions for future research, for those who wish to further investigate the manosphere.

6. Conclusion

To answer our first research question (How are gendered social actors represented across five manosphere communities on Reddit?), we can note the following: male and female social actors are mostly referred to in terms of their gender (e.g. *men*, *women*) and identified by employing romantic relational terms. Men in particular are denoted with kinship terms. As expected in view of previous studies, female social actors are more often derogated and ascribed negative sexual terms. While there are some instances of a counter-discourse, male and female social actors are mostly represented as homogenous, dichotomous groups. Nevertheless, a variety of personality types and possible changes are assigned to men.

Discussions around gender dynamics are integral to the manosphere: comparing men and women, commenters present the latter as enjoying social privileges, while men are represented as disadvantaged. Although women and men are both portrayed as perpetrators and victims of violence, women are specifically referred to as framing men for sexual violence. Despite the alleged disadvantages faced by men, they are still talked about as sexually, professionally and economically successful. Both female and male social actors are represented in terms of their physical appearance, with women and girls also described in terms of how sexually attractive commenters perceive them to be. Female social actors are moreover shown as passivated in sexual and romantic contexts, and are often ascribed descriptors and mental processes which denote their wishes, thoughts and feelings. This finding echoes much corpus linguistic and feminist critical discourse analysis work alike (e.g. Caldas-Coulthard and Moon, 2010; Mills, 2008).

With the common ground across the five subsections of the manosphere established, future research could interrogate relationships between users in comment threads using more qualitative methods. In this study, we did not have space to cover a larger number of gendered social actors key-keywords (see Lawson (forthcoming) for such a discussion), investigate the use of argumentation strategies and framing devices, or look at the representation of feminism and feminists, all of which would constitute fruitful directions for future research. In this study, we attempted to give a robust analysis of the usage of c-collocates for the four most commonly mentioned gendered social actors in our data, in order to address our research question on how gendered social actors are represented across the manosphere. Thus, to consider the sheer number of c-collocates this analysis generated, this necessitated a broad approach, whereas future research could choose one category of c-collocates to analyse in further detail.

In answer to our second research question (What ideologies of gender are social actor representations in the manosphere related to?), commenters show a strong belief that there are two clearly delineated and diametrically opposed genders. While their socio-cognitive representation of men is somewhat differentiated, they seem to believe that women in particular are a homogenous group with many negative traits. Manosphere discourse further reflects a belief that relations between genders are characterised by an imbalance of sexual and economic power, giving rise to an expectation that women's actions will put men at a disadvantage. While the ideology of the manosphere is characterised by heteronormativity,

commenters seem to see the current state of gender relations as violating their norms of appropriate behaviour in women.

The overall conceptualisations of female social actors across the manosphere can be seen as an extension of mainstream representations in general corpora of English (see e.g. Romaine, 2000; Pearce, 2008; Taylor, 2013). Given the “incremental effect of discourse” (Baker, 2006:13), it is likely that interaction with and within manosphere communities reinforces and amplifies ideological beliefs about gender, potentially radicalising members. What is more, the more extreme beliefs of the manosphere may be visible across other online communities such as gaming communities and alt-right spaces (see Massanari and Chess, 2018), which indicates a degree of mainstreaming. Future work will have to identify such mainstreaming and raise awareness about the risks associated with it.

References

- Baker, P. 2004a. ‘Unnatural acts’: Discourses of homosexuality within the House of Lords debates on gay male law reform. *Sociolinguistics* 8(1), pp 88-106.
- Baker, P. 2004b. Querying keywords: Questions of difference, frequency, and sense in keywords analysis. *Journal of English Linguistics* 32(4), pp 346-359.
- Baker, P. 2006. *Using Corpora in Discourse Analysis*. London: Continuum.
- Baker, P. 2014. *Using Corpora to Analyze Gender*. London: Bloomsbury.
- Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyzanowski, M., McEnery T., & Wodak, R. 2008. A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK Press. *Discourse & Society* 19(3), pp 273-306.
- BBC. 2019. ‘Pick-up artist’ Adnan Ahmed jailed for two years for threatening behaviour. *BBC* [online]. 22 Oct. Available at: <https://www.bbc.com/news/uk-scotland-glasgow-west-50137960> [Accessed 11 Dec 2020]
- Brezina, V., Timperley, M., & McEnery, T. 2018. #LancsBox v.5.1.2 [software]. Available at: <http://corpora.lancs.ac.uk/lancsbox/> [Accessed 22 December 2020]
- Caldas-Coulthard, C., & Moon, R. 2010. ‘Curvy, hunky, kinky’: Using corpora as tools for Critical Discourse Analysis. *Discourse & Society* 21(2), pp 99-133.
- Cecco, L. 2020. Canada police say machete killing was ‘incel’ terror attack. *The Guardian* [online]. 19 May. Available at: <https://www.theguardian.com/world/2020/may/19/toronto-attack-incel-terrorism-canada-police> [Accessed 04 Dec 2020]
- Clark, H. 1996. *Using Language*. Cambridge: Cambridge University Press.
- Connell, R., & Messerschmidt, W. 2005. Hegemonic masculinity: Rethinking the concept. *Gender & Society*, 19(6), pp 429-459.
- Davies, M. 2010. *Corpus of Contemporary American English* [online]. Available at: <https://www.english-corpora.org/coca/> [Accessed 11 December 2020]
- Dayter, D., & Rüdiger, S. 2016. Reporting from the field: The narrative reconstruction of experience in pick-up artist online communities. *Open Linguistics* 2(1), pp 337-351.
- Durrant, P., & Doherty, A. 2010. Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory* 6(2), pp 125-155.
- Egbert, J., & Biber, D. 2019. Incorporating text dispersion into keyword analyses. *Corpora* 14(1), pp 77-104.

- franzke, a. s., Bechmann, A., Zimmer, M., Ess, C. & the Association of Internet Researchers. 2020. *Internet Research: Ethical Guidelines 3.0*. Available at <https://aoir.org/reports/ethics3.pdf> [Last accessed 12 Apr 2021]
- Gabrielatos, C., & Baker, P. 2008. Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK press, 1996-2005. *Journal of English Linguistics* 36(1), pp 5-38.
- Ging, D. 2017. Alphas, betas, and incels: Theorizing the masculinities of the manosphere. *Men and Masculinities* 22(4), pp 638-657.
- Heritage, F., & [Author]. 2020. Incels, in-groups, and ideologies: The representation of gendered social actors in a sexuality-based online community. *Journal of Language and Sexuality* 9(2), pp 153-180.
- Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Jane, E. 2018. Systemic misogyny exposed: Translating Rapeglish from the manosphere with a random rape threat generator. *International Journal of Culture Studies* 21(6), pp 661-680.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, Y. 2014. The Sketch Engine: Ten years on. *Lexicography* 1(1), pp 7-36.
- Author. 2014. Applying social cognition research to critical discourse studies: The case of collective identities. In: C. Hart and P. Cap (eds). *Contemporary Critical Discourse Studies*. London: Bloomsbury, pp. 147-165.
- Author. 2020. The men and women, guys and girls of the ‘manosphere’: A corpus-assisted discourse approach. *Discourse & Society* 31(6), pp 607-630.
- Lawson, R. (forthcoming). *The Linguistic Construction of Contemporary Masculinities: Cultures, contexts, constraints*. Oxford: Oxford University Press.
- Lawson, R., & [Author]. 2017. “You need to become prime dick”: A corpus-based analysis of self-help discourses and gender constructions in online seduction communities. 116th annual American Anthropological Association meeting, Washington D.C., 30 November 2017.
- Lewis, H. 2019. To Learn About the Far-Right, Start with the ‘Manosphere’. *The Atlantic* [online]. 07 Aug. Available at: <https://www.theatlantic.com/international/archive/2019/08/anti-feminism-gateway-far-right/595642/> [Accessed 04 December 2020]
- Lumsden, K. 2019. “‘I want to kill you in front of your children’ is not a threat. It’s an expression of a desire, not of an intent”: Discourses of trolling and gendered violence on r/MensRights. In: K. Lumsden and E. Harmer (eds) *Online Othering: Exploring violence and discrimination on the Web*. Basingstoke: Palgrave Macmillan, pp. 91-115.
- Mackenzie, J. 2017. Identifying informational norms in Mumsnet Talk: A reflexive-linguistic approach to internet research ethics. *Applied Linguistics Review* 8(2-3), pp 293-314.
- Marwick, A. E., & Caplan, R. 2018. Drinking male tears: Language, the manosphere, and networked harassment. *Feminist Media Studies* 18(4), pp 543-559.
- Martin, J., & White, P. 2005. *The Language of Evaluation: Appraisal in English*. New York: Palgrave Macmillan.
- Massanari, A., & Chess, S. 2018. Attack of the 50-foot social justice warrior: the discursive representation of SJW memes as the monstrous feminine. *Feminist Media Studies* 18(4), pp 525-542.
- Mills, S. 2008. *Language and Sexism*. Cambridge: Cambridge University Press.
- Moon, R. 2014. From gorgeous to grumpy: Adjectives, age and gender. *Gender and Sexuality* 8(1), pp 5-41.
- Pearce, M. 2008. Investigating the collocational behaviour of MAN and WOMAN in the BNC using SketchEngine. *Corpora* 3(1), pp 1-29.

- Potts, A., & Baker, P. 2012. Does semantic tagging identify cultural change in British and American English? *International Journal of Corpus Linguistics* 17(3), pp 295-332.
- PRAW. 2020. *PRAW: The Python Reddit API Wrapper* [online]. Available at: <https://praw.readthedocs.io/en/latest/> [Accessed 11 December 2020]
- Rayson, P. 2008. Log-likelihood calculator. UCREL web server. Available at: <http://ucrel.lancs.ac.uk/llwizard.html> [Accessed 08 April 2021]
- Reddit. 2016. Reddit API terms of use. Available at: <https://docs.google.com/forms/d/e/1FAIpQLSezNdDNK1-P8mspSbmtC2r86Ee9ZRbC66u929cG2GX0T9UMyw/viewform> [Accessed 12 Apr 2021]
- Reddit. 2021. Reddit user agreement. Available at: <https://www.redditinc.com/policies/user-agreement> [Accessed 12 Apr 2021]
- Ribeiro, M., Blackburn, J., Bradlyn, B., De Cristofaro, E., Stringhini, G., Long, S., Greenberg, S., & Zannettou, S. 2020. *From Pick-Up Artists to Incels: A Data-Driven Sketch of the Manosphere* [online]. Available at: https://www.researchgate.net/publication/338737324_From_Pick-Up_Artists_to_Incels_A_Data-Driven_Sketch_of_the_Manosphere [Accessed 04 December 2020].
- Rodger, E. 2014. *My Twisted World: The story of Elliot Rodger* [online]. Available at: <https://assets.documentcloud.org/documents/1173619/rodger-manifesto.pdf> [Accessed 17 December 2020].
- Romaine, S. 2000. *Language in Society: An introduction to sociolinguistics* (2nd ed). Oxford: Oxford University Press.
- Rüdiger, S., & Dayter, D. 2017. The ethics of researching unlikable subjects. *Applied Linguistics Review* 8(2-3), pp. 251-269.
- Scott, M. 1996. WordSmith Tools Help. Oxford: Oxford University Press.
- Scott, M. 1997. PC analysis of key words - and key key words. *System* 25(2), pp 233-245.
- Scott, M. 2016. WordSmith Tools 7 [software]. Available at: <https://www.lexically.net/wordsmith/> [Accessed 22 December 2020]
- Sigley, R., & Holmes, J. 2002. Looking at girls in corpora of English. *Journal of English Linguistics* 30(2), pp 138-157.
- Slobe, T. 2018. Style, stance, and social meaning in mock white girl. *Language in Society* 47(1), pp 541-567.
- Stubbs, M. 1996. *Text and Corpus Analysis*. Oxford: Blackwell.
- Taylor, C. 2013. Searching for similarity using corpus-assisted discourse studies. *Corpora* 8(1), pp 81-113.
- van Leeuwen, T. 2008. *Discourse and Practice: New tools for Critical Analysis*. Oxford: Oxford University Press.
- Van Valkenburgh, S. 2018. Digesting the red pill: Masculinity and neoliberalism in the manosphere. *Men and Masculinities*, 1-20. doi:[10.1177/1097184X18816118](https://doi.org/10.1177/1097184X18816118)
- WebCorp Mini-Web Corpus. 2010. *WebCorp Linguist's Search Engine*. Available at: <http://wse1.webcorp.org.uk/home/syn.html> [Accessed 11 December 2020]
- Wright, D. 2020. The discursive construction of resistance to sex in an online community. *Discourse, Context & Media* 36(1), pp 1-9.