Illusory Correlations: An Investigation in a Social Media Setting

Abeba Birhane, Róisín Fitzpatrick, Caoimhe Kavanagh, John Rogers

Computer Science and Informatics, University College Dublin, Ireland
abeba.birhane@ucdconnect.ie, roz2030@hotmail.com, Caoimhe.Kavanagh@ucdconnect.ie,
John.Rogers.1@ucdconnect.ie

Abstract. This experiment investigated whether illusory correlations (ICs) exist within social media settings, and the likelihood of the occurrence of marked differences between individuals in their IC rates. Eighty individuals participated in the study, with 71 used for analysis. The experiment consisted of three tasks: a learning phase and two test phases, A and B. Participants were first presented with Facebook statuses from two groups, Purple or Orange, followed by questions on various aspects in phases A and B. Statistical analysis of the results showed no significant difference between the groups, however further analysis of subdivided groups showed higher IC rates in some groups. Confidence levels were found to correlate with the emergence of IC. Additionally, the results demonstrated the emergence of a 'reverse' IC that has been previously unreported in the literature. Qualitative analysis reflected a significant overall IC between the groups.

1 Introduction

Social discrimination or stereotyping has substantial and far reaching implications for society, and understanding the cognitive aspects that give rise to this phenomenon is crucial. In any society we would expect desirable and pro-social behaviours to be the norm, whereas behaviours on the opposite ends of those spectrums the exceptions. A problem arises when we realise that individuals, without evidence to the contrary, have a cognitive-based tendency [4] to attribute more frequently occurring behaviours with majority groups and less frequent behaviours with minority groups. This tendency where an individual perceives that a statistical relationship exists between two variables, where in fact none exists, is referred to as an illusory correlation (IC) [2]. This phenomenon is exemplified by an instance in which a false association is formed between generally uncommon negativebehaviours and a minority group – contributing to the formation of stereotypes [7]. Traditional social cognition approachs argues that we form stereotypes by means of our cognitive processes which rely on strategies such as the availability heuristic [7]. The term illusory correlation, first coined by Chapman in his 1967 landmark paper, refers to the systematic errors which are commonly made in the estimated correlation between two classes of events [2]. ICs were then first demonstrated empirically by Hamilton and Gifford in 1976, along with the first attempts at modelling the phenomenon. Despite the fact that the

IC phenomenon has been shown to be robust [3], [4], [9], there has not been a satisfactory explanation as to why it occurs or the cognitive mechanisms behind it. Nonetheless, a number of recognised, yet conflicting theories have been offered as explanations for this phenomenon.

The distinctiveness-based explanation (DBE), first proposed by [4] has been one of the most dominant theories to account for ICs. According to this explanation, an individual will estimate a higher rate of co-occurrence between two events when the events in question are unusual, rarely occurring, or are otherwise salient. This model proposes that the increased salience of the events in question leads to information concerning them being memorised and then recalled more easily, with a higher assumption of co-occurrence. Ethnic or social minorities are by definition less frequently occurring in a population than the respective majorities, and negative or anti-social behaviour is typically perceived as deviant and less prevalent than pro-social behaviour. The DBE asserts that two distinct pieces of information become implicitly linked during the encoding phase, resulting in the recall of either piece of information involving the retrieval of the paired 'unit' of events [9]. However, empirical evidence for distinctiveness based IC seems scarce. According to [6], there is no direct evidence for the idea that we have enhanced memory as a result of biased learning of infrequent pairings. Furthermore, the indirect evidence does not stand up to scrutiny.

Memory for rare events was not enhanced, and if anything it was shown to be impaired [6] according to their incomplete learning account (ILA). This account proposes that the assumption of selective memory or bias is unnecessary. The ILA instead argues that there is an increased number of learning instances for both the majority group and the positive characteristics, which results in these pieces of information being learnt more thoroughly. As a result the relative lack of information concerning the minority group and negative characteristics prevent the observer from learning the frequency of co-occurrence accurately, resulting in the over-estimations known as Ics [6].

A connectionist approach which uses autoassociative networks that mimic learning processes proposed by [10] proposes that observers construct a mental representation which consists of connections between social groups, episodic behavioural information, and an overall evaluative impression. Observers are considered to gradually develop prototypes of both majority and minority groups that include not only evaluative information, such as positive or negative feelings, but also store specific behavioural information. With every encounter the prototype is updated via a connectionist learning algorithm. In this way information concerning social groups and their behaviours are encoded twice: first as an overall impression of the group, and second as a specific episodic memory about the group or its members' behaviour. As the connections between the majority social group and overall impression are strengthened, the connections between these events and the other areas are weakened: those being the majority group and episodic memories, and the overall impression and minority group. Accordingly, this shift of potency results in the connections between the two smaller groups (minority event and episodic memory) being unaffected and remaining intact. ICs are then formed as an unintentional by-product of the cognitive mechanisms updating the mental representations [10].

On the other hand [3] have argued that IC is effective at all stages of information processing and not a product or by-product of a single stage of cognitive functioning. Following their examination of ICs between person types (a student versus a clerk) and educational attitudes (liberal versus authoritarian) under conditions where selective recall processes are highly unlikely, they argued that ICs are effective when the stimulus is perceived and encoded; when information is constructed on cue recall tests; when statistical properties are estimated; and when the impressions of the target persons are expressed within an adjective space. They found that there was a marked difference between individuals in their ICs and that the group effect was due to the overly biased response of a dozen participants. This implies that ICs, even though observed universally, may not be inherently dictated by the laws of cognitive processing, but rather are a product of arbitrarily chosen or learned strategies.

The concept of using social media, Facebook (FB) specifically, as a medium to investigate ICs is quite novel. In a recent study [6] have used Facebook-like personal profiles to investigate the mechanisms behind ICs in the context of attention shift in category learning. They proposed an inverse base-rate effect (IBRE) as a mechanism by which IC arises when frequent categories are paired with corresponding common attributes and infrequent categories with corresponding rare attributes. However, no evidence was found for IBRE as an explanation of ICs: although FB was used in this study, IC was not directly investigated.

1.1 The Current Experiment

In contemporary society, a large amount of the social information that an individual gains originates from a social media platform. The aim of the current study was to investigate whether ICs can also be found in individuals perceptions of FB statuses. The formation of ICs on social media is of great importance and relevance as an increasing amount of our social interaction takes place online [5]. Previous literature has shown that the IC phenomenon has been demonstrated to be robust across a variety of methods and in a variety of situations.

One of the main aims of the experiment was to investigate whether ICs will form based on the evaluative information that is present in FB statuses. Given that ICs have been implicated in a variety of social-cognitive settings (e.g. stereotyping), it is appropriate that empirical investigation takes place in a social media setting. Additionally, with the rising interest in social media among scientists, the current experiment adds to a rapidly building body of literature in this area.

Finally, previous studies in ICs tend to use vignettes rather than information presented in a more ecologically valid setting, and the current study aims to test the strength of the phenomenon outside of the lab. Given previous research suggests that generally, the majority group is judged more favourably compared to minority group, the current experiment consists of the following hypotheses: 1) Desirability and likeability scores for the majority group will be significantly higher than the minority group's scores for positive test items. 2) Negative test items will be disproportionately assigned to the minority group. 3) Likeability scores for the neutral statuses will be significantly higher when they are presented as made by the majority group than they were originally rated. We also further investigated [3]'s claim that there are marked differences between individuals in their IC rates.

2 Method

2.1 Participants

Eighty participants took part and nine were excluded from analysis due to incomplete responses, resulting in a final N = 71 (32 males and 39 females). Participants were recruited by convenience and referral sampling, predominantly through FB. Since FB status updates were used as a medium for inducing ICs, it was a necessary criterion that all participants were active FB users. The majority of participants (54%) fell in the age range 18-25, with a further 21% between 26 and 35 and the remaining 24% being 36+. The majority of participants reported active FB usage, with 68% reporting that they use the site daily, and only 8% reporting they use the site once or twice a week. The remaining 24% used FB between two to five times per week.

2.2 Materials

2.2.1 Status Selection. The status updates used in this experiment were gathered from FB Newsfeeds. These statuses were edited to remove identifiable information such as names of people or places as well as colloquial language in order to illuminate already formed biases associated with particular names or places. For example, 'I'm loving the Irish weather' was edited as 'I'm loving this weather' and the name and image of the person was replaced by a white FB template with Orange background (see Fig 1.). In order to avoid experimenter bias, the statuses were rated independently by 76 participants in a precursor ratings task by means of distributing these statuses online using Google Forms. Over 100 statuses were selected and distributed online through FB. The perceived desirability of the author of each status was recorded on a Likert-style five-point scale from extremely undesirable to extremely desirable. Desirability was defined as how positive or negative an impression the reader was left with of the author after reading their status. A neutral status was defined as one which made no impact on the reader, and would be scrolled past without a second thought.

The average rating of each status was then used to divide the statuses into four distinct groups: positive statuses were defined as those which fell between a score of 3.5 and 4.5, negative between 1.5 and 2.5, neutral between 2.5 and 3.5, and extreme statuses fell below 1.5 or above 4.5. Extreme statuses were excluded from further use, with the justification that if a status is regarded as too positive or negative, it becomes inherently memorable and would be unsuitable for this study [10].

2.2.2 Stimuli. FB newsfeed templates which resembled the real newsfeed were created and made anonymous by removing and censoring identifiable information, such as name, profile picture, and FB group membership. The statuses were then inserted into these templates and divided randomly between the majority and minority group: each group had twice as many positive statuses as they had negative, with the majority group having twice as many statuses overall as the minority group (20:10, 10:5 status ratio of positive: negative for each group). Consequently there were 45 total status images for the learning phase of this project. The two groups were distinguished using differently coloured profile pictures, with orange representing the majority group and purple representing the minority group (Fig 1). Participants were

unaware of any existing differences between the two groups. The aim of the coloured profile pictures was to eliminate possible pre-existing biases that a participant may have regarding names, sex, race, age, place of birth, schools attended, etc. However researchers also wanted to use somewhat established group dividers as the use of arbitrary group signifiers, such as 'Group A and Group B' has been seen to exaggerate IC effects [1]. Using a coloured but anonymous profile picture as well as individual but censored names for each status enables the researchers to control for these social biases, whilst still demonstrating to participants that each status was posted by a distinct person belonging to one of two relevant groups.

Fig. 1. *Stimuli used to denote group membership to the Purple or Orange group, respectively.*

The task consisted of three parts: the learning phase and two testing phases, hereafter referred to as test phase A and B. In the learning phase statuses were presented in a random order, featuring a group's icon to denote authorship. Participants were required to read each status and get impressions of the statuses associated with each group. In test phase A, a total number of 20 (ten positive and ten negative) statuses were presented without group denotations. Participants then predicted the group membership of that status's author; rated their confidence in their group assignment choice; their perceived desirability of the author upon reading the status; and finally how likely they would have been to 'like' that status had it appeared on their own Newsfeed. In test phase B participants were presented with ten statuses which were previously evaluated as neutral and non-impactful. These statuses featured an Orange (posted by a member of the majority group) profile picture. Participants again rated their likelihood to have 'liked' that status if they had seen it on their own Newsfeed.

2.2.3 Software. The project was built and presented in Python code using the program PsychoPy, [8]. This enabled the display of experimental statuses to appear exactly like real FB status updates while excluding some information like names that participants might have already formed biases on. All of the FB statuses and instruction pages were formatted in JPEG images. Participants performed the task on researchers' laptops. Prior to each phase participants were provided with clear instructions on the task. Each instruction screen included a five second minimum delay before the participant could move to the next screen by pressing the 'Enter' key: this prevented accidental skipping. The program then displayed each of the learning phase statuses in a randomised order, featuring a two second minimum delay before the participant could progress, for the same reason. The ratings for the test phases were recorded on a Likert-type scale. In order to familiarise participants with the experiment, they were presented with a practice phase prior to the test phases. After selecting their desired value, they had the option to change the value or confirm it via button click. The group assignment question was a two-point scale, while the other questions featured a five-point scale.

2.2.4 Other materials. Participants were provided with an information sheet describing the nature of the study and their task. Relevant information, such as age, gender, and the frequency of FB use were also recorded. We also had a question to be answered only after the task was completed: whether participants had a particular impression of the Purple and the Orange group upon reading such statuses.

2.3 Procedure

The task was conducted in residential environments. Participants were briefed and instructed on how to complete the various phases of the study. They were given the information sheet explaining what the study was concerned with, and what their involvement entailed and signed a consent form prior to their engagement. The task of the researcher was to initialise the experiment and give the participant space to complete the task. Upon completion of the task, participants were debriefed with further discussion of the subject of the experiment, and requested to fill out the demographic sheet. They were then thanked for their contribution.

2.4 Data analysis

Data was organised according to the group assignment made by participants in test phase A. As the statuses in test phase B were previously unseen, participants were using prototypes of the Orange and Purple group formed in the learning phase to assign groups to statuses The data was first analysed by examining the frequency of group assignment made to each test item: as each participant (N = 71) engaged in 20 test items (10 +ve and 10 -ve) in test phase A, a total of 1391 instances, excluding 29 missing due to technical difficulty, were used for this analysis. Subsequently, participants were separated into three subgroups for further analysis, which will be elaborated on below. Finally, qualitative analysis was carried out.

3 Results

3.1 Statistical

The analysis' point of departure was the desirability and likeability scores for each test item. During the initial testing phase, participants rated 20 test items consisting of equal numbers of positive and negative statuses (freq = 1390) as well as assigning them to a group. Assignment to the orange group occurred 51% of the time overall, 47% of the time for negative statuses and 55% for positive statuses. A battery of t-tests also showed that there were no significant differences in likability or desirability between Orange+ and Purple+ groups, or Orange- and Purple- groups. Thus no evidence was found for the first hypothesis.

Self-reported confidence measures were then controlled for two reasons. Firstly, as there was no objective difference between groups in terms of their positive and negative statuses (except presentation ratios) it is impossible for a participant to say, with any legitimate certainty, which group an unseen status, belongs to. It was expected that individuals with low confidence scores were more likely to engage in random guessing in their group assignments. Secondly, by taking into account individual dif-

ferences and how certain individuals may be particularly susceptible to IC, it was deduced that individuals with higher confidence levels may demonstrate a higher level of IC. High confidence responses were taken as those scoring either 4 or 5. An independent samples t-test examined differences in desirability for Orange- and Purple-, (of the high confidence responses) and found that Purple- (M = 1.72, SD = .95) scores were not significantly lower than Orange- (M =1.91, SD = 0.97) scores: t (290) = 1.63, p = 0.052.

An independent samples t-test showed that likelihood scores were significantly [t(296)= 2.49, p < 0.01] lower for Purple- (M = 1.39, SD = .92) than Orange- (M =1.68, SD =1.07). It was found that individuals who are confident in their group assignments were less likely to 'like' the negative statuses of individuals belonging to the minority group supporting the prediction that higher confidence is likely to be associated with higher level of IC.

For low confidence individuals, significant differences [t(112) = -2.94, p < 0.01] were found between likeability scores for Orange+ (M = 2.80, SD =1.45) and Purple+ (M = 3.44, SD = 1.47). Additionally low confidence individuals rated their likability of assigned Purple- statuses (M = 1.72, SD = 1.06) as higher than those of Orange- (M = 1.48, SD = 0.83); [t(188) = -1.66, p = 0.048). An analysis investigating the number of instances where individuals favoured selecting one group over the other for either positive or negative statuses was then conducted and is shown in table 1.

Table 1. *Frequency of participants leaning towards one group*

Individual preference	Orange+	Orange-	Purple+	Purple-
70%	27	22	24	31
80%	23	15	12	20

Following this, the probability of whether these individuals would assign Purple/Orange to either a positive or negative status in test stage A, was calculated. Individuals who rated the neutral statuses as positive in test phase B were 10% more likely to assign positive statuses to the Orange group in the test phase A. Likewise, individuals who voted the neutral statuses as negative were 10% more likely assign negative statuses to the Purple group.

When examining the complete dataset there were few statistically significant results. However upon closer inspection, it became clear that the results were not all falling into a neutral range, but rather the majority of responses were on opposite ends of the scale and were cancelling each other out (see Fig 2). After this initial analysis, it became evident that participants fell into one of three subgroups: participants that formed the IC as expected (Expected Illusory Correlation, hereafter ExpIC; n = 31), participants that displayed no preference or aversion to either group (No Illusory Correlation, hereafter NoIC; n = 16), and participants that formed the 'Unexpected' IC (Opposite Illusory Correlation, hereafter OpIC; n = 24). ExpIC were participants that assigned ≥70% of negative statuses to the Purple group, as hypothesised in this experiment and previous research. OpIC assigned ≥ 70% of negative statuses to the Orange group, which is a trend unprecedented in and unexpected by previous literature.
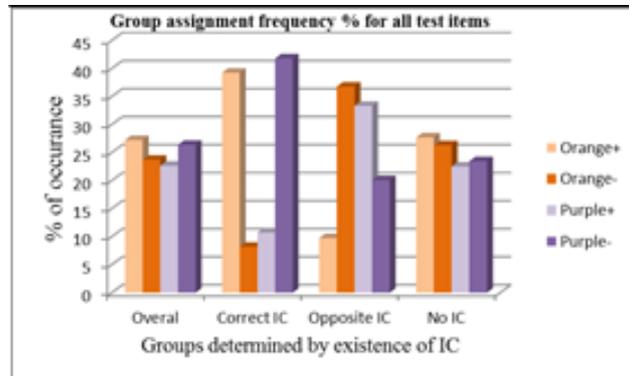
Fig 2. *Group assignment frequency % of all test items*

It appears that this subgroup formed a 'reverse' or 'opposite' IC, and this finding is discussed in detail below. Finally NoIC indiscriminately chose either Orange or Purple for negative test items, indicating that they formed no IC. The participants in this subgroup did not differ particularly from those in the other subgroups in terms of age, sex or ethnicity. This supported [3]'s assertion that there are individual differences in the susceptibility to IC formation and is discussed further below. An independent samples t-test found that ExpIC (M = 3.12, SD = 1.3) scored the neutral statuses significantly higher [t (414) = 2.43, p< 0.01] than OpIC (M = 2.83, SD = 1.25). NoIC did not differ significantly for neutral scores indicating that individuals who formed the Expected IC also demonstrated an exaggerated preference for statuses when they believed them to be authored by the Orange group.

Table 2. Differences between ExpIC, OpIC and NoIC for desirability and likeability for all test items.

			DesO+				LikO+			
Var 1	Var 2	Mean Var1	Mean Var2	T	Df	sig	Mean Var1	Mean Var2	T	df	Sig
ExpIC	OpIC	3.70	3.47	1.61	68	0.06	3.36	2.70	3.22	63	0.01**
OpIC	NoIC	3.47	3.28	1.09	112	0.14	2.70	3.02	-1.33	100	0.09
ExpIC	NoIC	3.70**	3.28**	3.31	188	0.01**	3.36*	3.02*	2.12	186	0.02*
			DesP+				LikP+			
Var 1	Var 2	Mean Var1	Mean Var2	T	Df	sig	Mean Var1	Mean Var2	T	df	Sig
ExpIC	OpIC	3.26**	3.77**	-2.94	73	0.01**	2.9**	3.56**	-3.06	74	0.01**
OpIC	NoIC	3.77*	3.47*	2.17	149	0.02*	3.56*	3.18*	2.31	170	0.02*
ExpIC	NoIC	3.26	3.47	-1.03	112	0.20	2.87	3.18	-1.24	100	0.15
			DesO-				LikO-			
Var 1	Var 2	Mean Var1	Mean Var2	T	Df	sig	Mean Var1	Mean Var2	T	df	Sig
ExpIC	OpIC	2.05*	1.88*	0.89	52	0.2*	1.90	1.88	1.42	51	0.08
OpIC	NoIC	1.88	1.97	-0.79	211	0.22	1.88**	1.52**	3.25	215	0.01**
ExpIC	NoIC	2.05	1.97	t Stat	64	0.35	1.3*	1.52**	1.82	59	0.04*
			DesP-				LikP-			
Var 1	Var 2	Mean Var1	Mean Var2	T	Df	sig	Mean Var1	Mean Var2	T	df	Sig
ExpIC	OpIC	1.80**	2.15**	-2.89	157	0.01**	1.49*	1.71*	-1.75	157	0.04*
OpIC	NoIC	2.15**	1.71**	2.84	193	0.01**	1.71	1.71	0.75	150	0.23
ExpIC	NoIC	1.80**	1.71**	-2.89	157	0.01**	1.49	1.71	0.75	150	0.23

A series of t-tests was then conducted in order to compare the differences between the three groups (Table 2.). The ExpIC group had significantly higher means for both desirability and likeability for Orange+ test items than both OpIC and NoIC groups. Additionally, the ExpIC group had a significantly higher mean than the NoIC group for both desirability and likeability for Purple- items. However, there was no significant difference between ExpIC and NoIC in relation to likeability scores for Purple-. For the OpIC groups, the mean for both desirability and likeability for Purple+ items were significantly higher than both NoIC and ExpIC groups. OpIC had a significantly higher mean for likability, but not desirability, than both ExpIC and NoIC groups for orange- test items. Contrary to previous findings, OpIC group rated positive purple items as being both more likeable and desirable than the other two groups.

3.2 Qualitative Analysis

Participant's qualitative responses, analysed in conjunction with their task responses, showed support for IC and in line with previous findings. Of the fifty-one participants that provided their characterisations of the groups, there was an overall tendency to favour the majority group. In characterising the Orange group, positive/desirable attributes were given 27 times, whereas undesirable/negative impressions were given 24 times. The Purple group however was ascribed with negative/undesirable characteristics in 34 incidents while only attributed 19 positive/desirable characteristics. Several participants characterised the Orange group as braggers, attention-seekers, over-sharers, or hormonal posters.

One common theme among the majority of participants was their general negative attitude towards FB, regardless of the experimental content or the group presented. This was reflected by the greater number (58) of undesirable/negative comments compared to (46) positive/desirable comments, made by participants when asked to provide their impression of each group. The most prevalent distinguisher between the Purple and Orange groups that was reported by participants was a believed difference in age. Although statuses were randomly divided between the groups during, participants felt the Orange group was older and more mature than the Purple group.

4 Discussions

4.1 Present Findings

Even-though the initial hypotheses had to be rejected, analysis of the results found considerable evidence to suggest that the ICs exist in social media and online interactions. Analysis of subgroups showed that ICs were evident. For instance, individuals who felt confident in their group assignments attributed not only a higher frequency of negative statuses to Purple but also relative lower/higher likeability and desirability scores for both the Purple- and Orange+ items. This suggests an IC had occurred. Additionally, low confidence individuals rated Purple- items as higher and Orange+ as lower; providing a possible insight into why OpIC may have occurred. Perhaps the core substance of the experiment comes from the separating of participants into three distinct groups. We hypothesised that after analysis, two distinct groups would

emerge (ExpIC, NoIC). An unexpected and substantial OpIC group mirrored the ExpIC group; nullifying their IC effect. Had our initial hypotheses been applied to each of the three groups, rather than conflating these distinct subgroups, our first hypothesis must be accepted for the ExpIC group; likewise, the opposite hypotheses would have been true for the OpIC group. What is evident from the comparison of the ExpIC and the NoIC group is that individual differences among participants can affect IC which is in line with [3]' findings that some individuals are more susceptible to ICs than others.

The OpIC group was an unexpected finding, with no previous literature discovering an opposite IC in an IC experiment. Confidence appeared to be a correlational factor in that high confidence individuals were more likely to rate desirability higher for Purple- and lower for Orange+, and vice versa. The qualitative data gathered by members of this group seemed to elucidate some of the cause of this group's appearance: although the most substantial block of statuses shown was Orange+, the Orange- block was measurably higher than the Purple-. IC literature is based on the evidence that humans disregard this objective information and form erroneous opinions concerning minority groups and behaviours; however this did not occur with the OpIC group. This might suggests that specifically in a FB setting, quantity can matter more than quality: participants felt that the Orange was the more negative of the groups, even though this itself is still an IC.

Testing phase B also yielded significant results. ExpIC individuals rated the neutral statuses as positive, whereas OpIC individuals tended to vote for the neutral statuses as being more negative. Given the neutral statuses were presented as posted by the Orange group, we can deduce that the formation of ICs, whichever direction it took, may have influenced individuals perceived likeability of the statuses. IC research does not often investigate this inverse trend: while it is typical to examine an aversion to the minority group, most researchers do not consider if the participant also displays a preference to the majority group. By asking participants to rate neutral statuses assigned to the majority group we found that this converse preference existed in the expected direction for both ExpIC and OpIC, with the NoIC group continuing to evaluate them as neutral. This was supported by the probability estimates discussed previously. The potential implications of this inverse finding suggests that frequency of exposure to a minority and majority group can affect approval ratings of not only the minority group but also the majority group.

Taken together, the results suggest that individuals who were confident in their answers were more likely to rate Purple- items worse than negative items they deemed as belonging to the Orange group. Additionally individuals who were uncertain in their answers were more likely to attribute higher likeability scores to Purple groups. As the overall qualitative characterisation of the Purple group was more negative when compared to the Orange group, the present analysis supports previous IC research findings [4],[9].

Even-though there were more positive statuses posted by the Orange group, some participants indicated that this led to a negative impression of the Orange group as a whole, which contradicts previous findings. Several participants characterised the Orange group as braggers, attention-seekers, over-sharers, or hormonal posters. This

highlights a key difference in the design of this experiment when compared to less recent studies: whilst this information would usually have been conveyed in a vignette, the present study presented the same information as being posted directly from the individual. Accordingly, that which is usually seen as a positive characteristic can have negative connotations when self- presented in a FB setting.

4.2 Limitations and Future Research

This experiment was carried out mainly at residences limiting the variety within the sample of participants. Most participants had connections to the researches either through university or family and friends. Given that FB users is arguably one of the largest populations worldwide, with over 1.49 billion users and regularly being ranked in the top most trafficked sites on the internet, the population sample used in this study is unlikely to represent the general FB population. This experiment could be improved by conducting the study online: this affords the chance to gather data from participants from a more varied demographic background, likely to be more representative of the FB user population. Nonetheless, the findings from this research are valid within the context in which it was carried out.

As the experiment was using status updates, i.e. statements with values attached, it is inevitable that individual's values differ, despite the fact that the statuses were rated for a consensus regarding its inherent positivity or negativity. These statuses could potentially mean different things to different participants and this was evident occasionally during debriefing of participants. 'Likeability', for example, was subjective and dependent on the kind of relationship one has with the person issuing status. A FB user is likely to 'Like' a status such as 'I can't wait to go on my holiday' if it was posted by a close friend or family member; the likelihood for 'Liking' that same status is much lower when it is posted by a person that they barely knew. It is recommended that future research takes a note of this important aspect and incorporate it when designing experiments. Similarly, complacency with or negative opinions towards FB were reported by certain participants, which complicates the findings of this research.

After the statuses were assigned as positive, negative, or neutral, they were divided randomly between the Orange and Purple groups to the desired frequencies. For the most part, excluding references to one group being more positive or negative than the other, the characteristics assigned to the groups by participants were arbitrary and reflected their general opinion and not the content of the statuses themselves. The only somewhat common theme was a perceived difference in age between the Orange and Purple groups. While this was not apparent in the statuses, perhaps there were age cues which were implicit in the statuses that happened to divide between the groups after random assignment. This again is an aspect for future research which may be considered. During debriefing participants claimed that it was strange to see an Orange member post such a "Purple-typical" status, indicating that matching statuses may not have had a significant effect.

A few technical issues arose during this study. The slider range used for test phase A and B featured five points, which could have been lengthened for increased sensitivity. Furthermore, the demographic information gathered failed to capture any of

the individual differences which effected IC formation, or lack thereof, which occurred in this experiment and is proposed by [3]. More detailed demographic information may have shed light on the underlying grounds of these individual differences. Previous research has shown that using arbitrary (e.g. Group A/Group B) instead of real-world (e.g. Class of 1989/Class of 1992) groupings for IC research leads to exaggerated or heightened ICs [1]. This implies that when a participant is not given a distinguishing feature between two groups, they will search for one or perhaps even create one. This is a confounding factor when trying to specifically examine ICs as they form naturally. However, in using real names or profile pictures social biases are introduced. The design of this experiment endeavoured to control for both extraneous social biases and these categorisation processes by using coloured profile pictures and censored names. However through debriefing participants it became clear that many were still focused on trying to find a difference between the groups, rather than accepting that the groups were already distinguished. This could potentially have led to exaggerated IC effects, as is seen in other papers. Future research may wish to seek other means to differentiate between their groups while still controlling for social biases.

References

1. Berndsen, M., McGarty, C., Pligt, J., & Spears, R. (2001). Meaning-seeking in the illusory correlation paradigm: The active role of participants in the categorization process. *British Journal of Social Psychology*, *40*(2), 209-233.
2. Chapman, L. J. (1967). Illusory Correlation in Observational Report. *Journal of Verbal Learning and Verbal Behaviour, 6*(1), 151-155.
3. Fiedler, K., Hemmeter, U., & Hofmann, C. (1984). On the origin of illusory correlations. *European Journal of Social Psychology, 14*(2), 191-201.
3. Fiedler, K. (2000). Illusory correlations: A simple associative algorithm provides a convergent account of seemingly divergent paradigms. *Review of General Psychology, 4*(1), 25-58.
4. Hamilton, D., & Gifford, R. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgements. *Journal of Experimental Social Psychology, 12*(4), 392-407.
5. Hunsinger, J., & Senft, T. M. (2013). *The social media handbook*. Routledge.
6. Kutzner, F. L., & Fiedler, K. (2015). No correlation, no evidence for attention shift in category learning: Different mechanisms behind illusory correlations and the inverse base-rate effect. *Journal of Experimental Psychology: General*, *144*(1), 58-75.
7. Matlin, M. (2005). *Cognition* (6th ed.). New Jersey: Wiley.
8. Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*, *162*(1), 8-13.
9. Risen, J. L., Gilovich, T., & Dunning, D. (2007). One- shot illusory correlations and stereotype formation. *Personality and Social Psychology Bulletin, 33*(11), 1492-1502.
10. Van Rooy, D., Van Overwalle, F., & Vanhoomissen, T. (2013). Illusory correlation, group size and memory. *Journal of Experimental Social Psychology*, *49*(6), 1159-1167.