British Journal of Social Psychology (2011), 50, 74–98
© 2010 The British Psychological Society

The British Psychological Society

www.wileyonlinelibrary.com

Shaping stereotypical behaviour through the discussion of social stereotypes

Laura G. E. Smith[1]* and Tom Postmes[2]

[1]The University of Queensland, St Lucia, Australia
[2]The University of Groningen, The Netherlands

In two studies, we demonstrate that small group discussions change the extent to which an activated stereotype affects performance in a relevant domain. In Study 1, female participants were asked why men are (or are not) better than them at maths. They generated their answers individually or through group discussion, and their subsequent maths performance was highest when they collectively challenged the stereotype and lowest when they collectively affirmed the stereotype. When participants affirmed the stereotype through discussion, they used more theories which supported the validity of the stereotype, compared to the individual thought condition; and consensus mediated the effect of group discussion on performance (relative to individual rumination). In Study 2, male and female participants affirmed or challenged the stereotype in same-gender discussion groups. After affirming the stereotype, women's performance decreased relative to their baseline scores and men's performance was 'lifted'. In contrast, when they challenged the stereotype, there was no difference between the performance of men and women on the maths test. This pattern of effects was mediated by confidence in mathematical ability. The findings support the idea that topical small group discussions can, in the short term, differentially alter the impact that stereotypes have on performance.

Over 25 years ago, Cockcroft (1982) reported that the performance of girls in mathematics exams in the UK was significantly lower than that of boys, and that this trend was replicated in many cultures around the world. Although a recent study of 3,000 twin pairs resulted in the conclusion that there is no genetic reason why the performance of boys and girls at school should differ (Haworth, Dale, & Plomin, 2009), the trend re-emerged in the UK in 2009, with results for 16-year-olds at school for General Certificate of Secondary Education qualifications showing that more boys than girls achieved grades A*–C after coursework was removed from assessment – the qualification now entirely dependent on exam results (Curtis, 2009). Furthermore, a cross-national meta-analysis showed that males generally outperform females at mathematics

*Correspondence should be addressed to Dr Laura G. E. Smith, School of Business, The University of Queensland, St Lucia, Qld 4072, Australia (e-mail: l.smith@business.uq.edu.au).

DOI:10.1348/01446610X500340

and have greater confidence in their mathematical ability (Else-Quest, Hyde, & Linn, 2010).

Evidence such as that described above fuels the public debate around the stereotype that males are equipped to outperform females at mathematics (Rosenthal & Crisp, 2006). For example, in 2005, the president of Harvard University publicly speculated that a reason for the underrepresentation of women in science and engineering is because of 'a different availability of aptitude at the high end' (see Dar-Nimrod & Heine, 2006, p. 435). These comments sparked a public outcry, with a prolific response to be seen on on-line discussion boards to this day. Alarmingly, there is mounting evidence that awareness of this stereotype influences the performance of women in maths tests (Marx & Roman, 2002; Nosek, Banaji, & Greenwald, 2002; Schmader, 2002; Spencer, Steele, & Quinn, 1999). At the same time, in other domains, there is evidence that discussion can change stereotypes (e.g., Haslam, 1997) and profoundly affect behaviour (e.g., Lewin, 1953). Therefore, it appears that exposure to discussions about stereotypes may have the potential to affect the behaviour of the targets of stereotypes. After all, stereotypes are socially shared representations of social groups (Kashima, 2000; Lyons & Kashima, 2003), and the process of learning about them is largely through some form of communication (e.g., Karasawa, Asai, & Tanabe, 2007; Ruscher, 2001; van Dijk, 1987). Yet while debate about the stereotype clearly continues in society, research has not yet addressed the impact this discussion has on the targets of the stereotype. The present research was designed to address this issue by drawing on insights from the small group communication and stereotyping literatures. Specifically, we focused on the impact of discussing the stereotype on targeted individuals' mathematical performance.

Stereotype threat

The negative impact that activated stereotypes may have on the performance of stigmatized individuals is known as *stereotype threat* (e.g., Aronson *et al.*, 1999; Steele, 1997; Steele & Aronson, 1995). Stereotype threat is a robust effect, shown to affect performance in a variety of domains, including women and career choices (Davies, Spencer, Quinn, & Gerhardstein, 2002), race and academic performance (Aronson, Fried, & Good, 2002; Steele & Aronson, 1995), race and maths performance (Aronson *et al.*, 1999; Smith & White, 2002), and social class (Croizet & Claire, 1998), to name but a few.

Researchers over the past decade have suggested that there are various types of, sources of, and responses to stereotype threat (for a review, see Shapiro & Neuberg, 2007), and along with each of these is a selection of bespoke moderators and mediators. Indeed, it is unlikely that there is a single universal mechanism behind stereotype threat – instead, a series of social, cognitive, and physiological processes may interrelate to cause the phenomenon (Schmader, Johns, & Forbes, 2008). The effect of stereotypes is not universally negative: stereotypes may also improve the performance of individuals. Evidence has emerged that the negative stereotyping of an out-group can be utilized to enhance or 'lift' in-group performance (Walton & Cohen, 2003).

While stereotype threat has the potential to be a substantial problem for members of stigmatized groups, individuals have both intrinsic and extrinsic motivations for resisting or challenging stereotypes, and there are both implicit and explicit methods by which the affect of stereotypes can be can be moderated (e.g., Smith & White, 2002; for reviews, see Rosenthal & Crisp, 2006; Sherman *et al.*, 2008). For example, over time Aronson *et al.* (2002; Good, Aronson, & Inzlicht, 2003) were able to reduce the stereotype threat effect

by informing participants that intelligence was malleable and not fixed. The effect has also been reduced through using relevant positive role models (Marx & Roman, 2002), by informing participants that the stereotype does not apply in the current context (Ouwerkerk, de Gilder, & de Vries, 2000; Spencer *et al.*, 1999), by blurring intergroup boundaries (Rosenthal & Crisp, 2006) and through self-affirmation strategies (Martens, Johns, Greenberg, & Schimel, 2006).

It appears therefore that the effect of these stereotypes can be reduced by challenging individuals' perceptions. However, phenomena like stereotypes have their effect because they are collectively shared representations of reality (Kashima, 2000; Lyons & Kashima, 2003), validated by a proportion of society (Karasawa *et al.*, 2007), and maintained by a certain degree of societal consensus (Hardin & Higgins, 1996). Therefore, changing individuals' perceptions of the situation does not necessarily change the origin of the threat: the perception that the stereotype is valid in the eyes of others. A challenge by one individual would not change the negative content of the in-group stereotype *per se*; it would merely bypass the threat for the moment. Instead, the group must participate in collective action to challenge or invalidate the stereotype together (Haslam, Salvatore, Kessler, & Reicher, 2008). One important way in which group members can do this is through *group discussion*.

Stereotypes and group discussion

During group discussion, many different processes govern stereotype transmission (e.g., Kashima, 2000; Lyons & Kashima, 2003; Semin, 2008). For example, individuals will often tune the content of their communication to their audience (e.g., Clark & Murphy, 1982; Higgins, 1992). One robust effect is that discussing a stereotype leads to increased consensus within the group on the stereotypic traits (e.g., Haslam, 1997; Haslam, Oakes, Reynolds, & Turner, 1999; Haslam, Turner, Oakes, McGarty, & Reynolds, 1998; Haslam, Turner, Oakes, Reynolds, *et al.*, 1998; Thompson, Judd, & Park, 2000). Significantly, perceptions of this group consensus validate individuals' beliefs and ideologies (Correll & Park, 2005; Festinger, 1950, 1954; McGarty, Turner, Oakes, & Haslam, 1993), increasing individuals' confidence in the accuracy of the views (Baron *et al.*, 1996; Luus & Wells, 1994; Petty, Briñol, & Tormala, 2002). If group members agree upon and validate particular reasons why a particular stereotype has (or does not have) a legitimate basis, they may infer that the associated behaviours are (or are not) normal for their group. This in turn will dictate the extent to which group members' behaviour conforms to the stereotype. For example, if women discussed the stereotype that males are better at maths than they are, and the emergent consensus supported this stereotype, they may experience decreased confidence in their mathematical ability, and subsequently worse performance on a maths test. On the other hand, if women agreed together that the stereotype is invalid, or if there was a lack of consensus on this issue, the stereotype would be comparatively less likely to negatively affect their performance. Discussion would be more powerful in affecting stereotypic behaviour than individual cognition because it provides this consensus (or dissensus) information.

Therefore, a strategy for accepting or changing a stereotype within the group would be to reach a consensus about the validity (or lack of validity) of the stereotype through discussion with other group members. The consensus (or dissent) that emerges within the group through discussion will then guide subsequent behaviour (cf. Chong & Druckman, 2007; Turner & Killian, 1972) due to its validating (or invalidating) effect. Indeed, past research suggests that group members often use the views they perceive

to be shared within their group to inform their social behaviour (Abrams & Hogg, 1990; Haslam *et al.*, 1999; Reicher, Spears, & Postmes, 1995; Sechrist & Stangor, 2001; Turner, 1985). Indeed, collectively challenging (or using) established beliefs within a group is often used a means to achieve social change (for both low and high status groups; see for example, Craemer, 2008; Mallett, Huntsinger, & Sinclair, 2008; Stott & Drury, 2004; Subasic, Reynolds, & Turner, 2009; van Zomeren, Postmes, & Spears, 2008). However, the extent to which the process of stereotype threat is interdependent with the intra-group dynamic of targeted group members is an important yet relatively unexplored question (Carnaghi & Yzerbyt, 2007). Therefore, the current research was designed to test the consequences of discussion about the stereotype that males are better at mathematics than females on maths performance.

The current research

We conducted two studies that examined whether the effect of the gender stereotype on maths performance may be moderated through challenging or affirming the stereotype through small group discussion (relative to no communication) with same-gender peers. By combining insights from research into small group discussion and the literature on stereotype threat, our goal was to directly investigate the impact of this process (validating or invalidating views through group discussion), which appears to be fundamentally involved in the effect of stereotypes on behaviour. In doing so, we contribute to the literature on the role of shared views on the effect of stereotypes on behaviour (Karasawa *et al.*, 2007) and locate the origin of this process in perceivers' ability discuss stereotypic views within their referent group.

STUDY 1

In the first study, we asked female participants to either challenge or affirm the statement that women underperform at maths relative to men. Importantly, they did this either through individual thought, or through group discussion. Therefore, we were able to compare the consequences of group discussion about the gender stereotypes to those of individual thought, and establish the unique impact of group discussion on stereotype-relevant behaviour. Furthermore, in order to examine the extent to which participants' consensus on the stereotypes affected their subsequent maths performance in each condition, we analysed the degree of consensus on the key points raised by participants and compared these for content across conditions.

Hypotheses

We hypothesized broadly that (a) group discussion would result in significantly different and more consensual content of the stereotype relative to the individual thought conditions. Next, (b) we predicted an interaction between type of reflection and topic of reflection. Specifically, we hypothesized that participants who affirmed the stereotype would have lower maths scores than participants who challenged the stereotype, but this difference would be achieved through group discussion rather than individual cognition. Finally, (c) we predicted that this interaction would be mediated by the extent to which participants had consensus on the topic of reflection.

Method

Participants and design

Participants were 75 female first-year psychology undergraduates from a university in the UK, recruited during class time. The university had a selective admissions policy, with top-level academic entry requirements. The undergraduate admissions policy states that all applications are considered and evaluated on the basis of individual merit. Of undergraduate admissions to the university in 2006, 92% were domestic or European Union students. Non-White students formed 4.9% of admissions, and 26% attended fee-paying schools prior to entering university. The mean age was 19.74 years old ($SD = 3.80$, range = 24). The design was a 2 (Topic: affirm stereotype vs. challenge stereotype) $\times$ 2 (Reflection: group discussion vs. individual cognition) between-subjects factorial design. Participants were randomly allocated to the conditions, and within those to three-person groups ($N = 25$).

Materials and procedure

Consenting participants were informed that they were taking part in a study on mathematical ability, which involved a modified brainstorming task and a maths test. The experimenter then split participants into groups by randomly assigning each participant a group number. Participants were informed that first of all, they would be taking part in an exercise to help them generate their opinions. They were then asked to discuss in their groups or think about alone, either 'Why it is true that men are better than women at maths', thus affirming the stereotype; or 'Why it is not true men are better than women at maths', thus challenging the stereotype. To ensure that participants in all conditions worked to a similar goal, they were all asked to write down their top five ideas. After 10 minutes, participants were asked to stop discussing or thinking about this. They then individually completed a maths test, which began with the following feedback: 'There is a stereotype in society that women are worse at maths than men. Gender differences have been found in the past on this maths test. This test is diagnostic of your numerical ability, so please do your best to answer the following questions'. The test contained four questions of approximately high school standard, which was to be completed in silence, individually and without the use of a calculator. Although there was no time limit on completing the test, the amount of time taken was recorded in order that its effect could be controlled. Participants were then debriefed.

Dependent measures

Maths performance

The performance measure was the total score on maths test, which contained four items: '3/5–1/4 = ?'; 'Make a the subject of this formula: $r = \sqrt{(4a + b)}$'; 'Write down the gradient and the intercept of the straight line graph: $2y = 4x + 8$'; '$2/3 + 2/4 = ?$'; and 'A formula is given by $y = mx + c$. Find the value of x when $y = 13$, $m = 5$ and $c = 3$'. Answers were marked either correct (1) or incorrect (0), and correct answers were totalled to give a score out of 4.

Consensus

We conducted a content analysis of the ideas listed by each participant. From this analysis, scores were calculated to represent the amount of consensus surrounding the content

of the stereotype in each group of three participants. To do this, first we calculated the average proportion that each category code was mentioned (Tables 1 and 2). Then, we computed how frequently each group of $N = 3$ participants mentioned each code. Next, for each group, we computed a relative score of how frequently they mentioned each code, minus the average proportion that the categories were mentioned in the condition. We then normalized the scores for each group, so that scores range from 0 to 1 (where 1 = complete consensus that a code either applied or did not apply).[1] The scores for each group were then aggregated across codes, to create one number per group, ranging from 0 (no deviation from average proportions) to 1 (consistent consensus).

Analytic strategy

Stereotype content

A content analysis was performed on the written description of the in-group and out-group stereotype provided by participants across conditions. The primary coder followed the iterative inductive and deductive coding procedures, recommended by Miles and Huberman (1994). In accordance with these procedures, the primary coder was not blind to the hypotheses. A second blind coder independently coded the data. Inter-rater agreement was 99% for the affirm condition and 95% for the challenge condition, and reliability was good, Cohen's (1960) $\kappa = .68$–1.00. Please refer to Tables 1 and 2 for definitions of the codes. The purpose of this analysis was to explore the content of the stereotype formed through group discussion or individual reflection for any differences. Therefore, the mean frequency with which the codes were mentioned in each condition was compared using t tests for responses in the affirm and contest conditions, respectively. In order to compare consensus on the stereotype across conditions, the consensus scores described above were analysed via analysis of variance (ANOVA) at the group level. Subsequently, in order to test whether group consensus mediated the effect of condition on maths score, we performed an analysis of covariance on group-level consensus and maths scores.

Maths performance

The responses to the measures were shaped by the small group interaction, therefore these data are interdependent. The intra-class correlation confirmed this, ICC = .31, indicating that 31% of the variability in maths scores was associated with differences between small groups. Multi-level modelling is capable of partialling out the variance caused by both group-level and individual-level factors (Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). Therefore, we conducted hierarchical linear modelling (HLM) analyses in HLM for Windows 6.03 (Raudenbush & Bryk, 2002) in order to analyse condition differences in maths scores. Multi-level analyses are appropriate for data in which one level of analysis (individuals, level 1) is nested within another (groups, level 2). This procedure examined the behaviour of the level 1 outcome (maths performance) as a function of level 2 predictors (condition). Dummy and contrast variables were created to represent the independent variables (i.e., a multi-level equivalent of the General Linear

[1] *In order for a score of '1' to represent consensus, if the category was mentioned by all members of a group, scores were divided by 1-condition average. If the category was not mentioned by all members of the group, scores were divided by the condition average itself.*

Table 1. Frequency of codes and code definitions for responses to the question, 'Why is it true that men are better than women at maths?' in the group discussion ($N = 21$) and individual thought ($N = 19$) conditions

Focus	Code		Definition	Example quotes	Group discussion	Individual thought
1. Traits	1.1	SPAT	Men have better spatial awareness than women	'Men are considered better at spatial awareness – better able to visualise'	38_a	32_a
	1.2	SIN	Men are better than women at (focusing on) single tasks	'Men have better focus on one task'	10_a	0_a
	1.3	SAT	Women have shorter attention spans than men	'Women have shorter attention spans'	$0_a^{\dagger}$	$16_b^{\dagger}$
	1.4	LOG	Men are more logical/factual/rational than women	'Men are more logically minded'	67_a	63_a
	1.5	NUM	Men are more numerically intelligent than women	'Men are better with numbers'	10_a	47_b
	1.6	BET	Women are better (than men) at things other than maths	'Women are… better at other things'	5_a	0_a
	1.7	CRE	Women are more creative than men	'Women are more creative'	$5_a^{\dagger}$	$26_b^{\dagger}$
	1.8	EMO	Women are more emotional (and therefore more irrational) than men	'In tests men don't panic as much as women so are better at the methodological approach of maths' 'Men do not think about feelings as much and so find it easier to think straightforwardly'	0_a	11_a
	1.9	VERB	Women are more verbally intelligent than men	'Women are better with words'	$10_a^{\dagger}$	$32_b^{\dagger}$
2. Behaviours	2.1	OTH	Women have more important, other things to think about/do	'Women have more important things to think about'	10_a	5_a
	2.2	ED	Men in particular prefer maths as an academic subject or are more likely to advance in their maths education than women	'Men are pushed harder in maths due to "manlier" jobs being more maths based'	48_a	37_a

Table 1. (*Continued*)

Focus	Code		Definition	Example quotes	Group discussion	Individual thought
				'Traditionally men were better educated, women expected to stay at home'		
	2.3	COMP	Men are more competitive than women	'Men are more competitive – like solving problems etc'	10_a	11_a
3. Roles	3.1	PROF	Maths is involved in male professions	'[Men] have more professional jobs like engineers which require advanced mathematical skills'	62_a	42_a
4. Theories	4.1	SOC	Societal/cultural cause e.g., gender stereotyping	'Perception of a male dominated world can lead to women lacking the chance to practice their mathematical skills'	62_a	58_a
				'Cultural assumption that men are better at maths than women'		
	4.2	BRA	Physical cause, i.e., differences in male and female brain function	'Men and women use parts of the brain differently. Maybe their brains work more laterally'	86_a	47_b
				'The area of the brain that deals with maths might be more developed in men'		
	4.3	EVO	Evolutionary cause	'Evolution – Men have had to use maths skills e.g., angles more than women'	$43_a^{\dagger}$	$16_b^{\dagger}$
5. Denial	5.1	EQ	Maths ability of men and women is equal	'Women aren't actually worse it's often seen like that'	0_a	19_a

Note. Percentages were computed at the individual level. Frequencies on the same row with different subscripts differ at $p < .05$; $^{\dagger}p < .10$.

Table 2. Frequency of codes and code descriptions for responses to the question, 'Why is it not true that men are better than women at maths?' in the group discussion ($N = 18$) and individual thought conditions ($N = 18$)

Focus	Code		Definition	Example quotes	Group discussion	Individual thought
1. Traits	1.1	ATT	Women are more attentive than men	'Women can concentrate better' 'Women have a longer attention span' 'On average, women tend to have longer concentration levels than men so tend to get better grades in maths'	6_a	11_a
	1.2	INT	Women are more (numerically) intelligent than men	'Women are more intellectual'	17_a	22_a
2. Behaviours	2.1	EX	Women do better in exams/at school than men	'Girls do better at school' 'Women get better exam results'	72_a	17_b
3. Roles	3.1	FEM	Women use maths professionally as well as men	'There are more female maths teachers' 'The number of women in maths related jobs in increasing' 'In a family environment women tend to do shopping and bills more than men so have to use maths more often with budgeting and have more practice'	22_a	28_a
	3.2	ROL	There are famous female role models who are good at maths	'Carol Vorderman!'	28_a	33_a
4. Theories	4.1	SOC	The gender difference is based on social expectation/stereotypes	'Social prejudices & opportunities may encourage different sexes to go for different subjects etc' 'Women have always been equally as good at maths but haven't had the chance to use it due to a male dominated society'	83_a	35_b
	4.2	NUR	Intelligence is determined by nurture, not nature (i.e., environmental factors, not gender)	'Environment determines your intelligence not gender'	33_a	6_b
5. Denial	5.1	EQ	Maths ability of men and women is equal	'Girls do just as well as boys at maths at school' 'Women are just as intelligent as men'	50_a	56_a

Table 2. (*Continued*)

Focus	Code		Definition	Example quotes	Group discussion	Individual thought
	5.2	DED	Deny that maths education favours men	'Taught the same. No different teaching determined by sex'	33_a	33_a
	5.3	DBRA	Deny that there is a physical difference between genders which affects mathematical ability	'No biological reason – no difference in brain activity'	22_a	17_a
				'In evolution terms women are now moving into male roles so the brain capacity would have to equalise'		
	5.4	DEVO	Deny that men have an evolutionary advantage	'There is no evolutionary advantage to men being better so they aren't'	11_a	11_a
	5.5	RST	Resistance to the stereotype	'Can't judge all men and women together'	33_a	50_a
	5.6	IND	There is individual variation in mathematical ability, not group (gender) variation	'It is down to the individual not gender of the person as to their maths ability'	17_a	17_a

Note. Percentages were computed at the individual level. Frequencies on the same row with different subscripts differ at $p < .05$.

Model approach). The estimated models were:

$$\text{Level 1 Model}: \quad Y = \beta_0 + R$$

$$\text{Level 2 Model}: \quad \beta_0 = \gamma_{00} + \gamma_{01} \times (\text{contrast1}) + \gamma_{02} \times (\text{contrast2}) + \gamma_{03} \times (\text{contrast3}) + U_0.$$

In these models, Y represents the dependent variable, R and U the errors at the individual and group level, respectively, and β and γ are regression coefficients at the individual and group levels, respectively. In the Results section, γ and p values are reported as a direct test of the relation between the independent (via contrast variables) and dependent variables. Means and standard deviations are reported in Table 3.

Table 3. Means and standard deviations of dependent variables, Study 1

Topic	Affirm		Contest	
Reflection condition	Discussion	Individual rumination	Discussion	Individual rumination
Maths test score				
M	1.40_b	2.21_a	2.78_a	2.06_{ab}
SD	1.05	1.03	1.06	0.94
Consensus score				
M	$.95_a$	$.78_b$	$.86_b$	$.76_b$
SD	.05	.05	.12	.02

Note. Consensus scores were calculated at the group level, maths scores were calculated at the individual level. Means in rows with different subscripts differ at $p < .05$.

Main effects and the two-way interaction were tested by means of dummy variables (Judd, McClelland, & Culhane, 1995; Rosenthal & Rosnow, 1985; Wilkinson *et al.*, 1999), and contrast variables were created to compare specific cells, following recommendations for contrast analysis in the testing of regression models by Cohen, Cohen, West, and Aiken (2003).

Results[2]

Stereotype content

In the affirm stereotype condition, participants listed an average of 4.65 ideas to answer the question, 'Why is it *true* that men are better than women at maths?' There were 17 distinct ideas in total. These were collapsed into five categories: traits, behaviours, roles, theories, and denial (see Table 1 for a full summary of the data).

Participants in the individual rumination condition more frequently mentioned traits which confirmed the gender stereotypes $t(38) = 0.19$, $p = .03$, with an average of 2.26 traits mentioned after individual thought ($SD = 1.41$), compared to an average of 1.43 traits mentioned in the group discussion condition ($SD = 0.98$). Examination of the

[2] *Space constraints prevent us from reporting all dependent variables here. Both studies measured social identification; however, we found no significant condition effects or significant moderation or mediation effects for this variable. Therefore, changes in identification could not explain the differences between conditions.*

codes mentioned within this category showed that participants in the individual thought condition were significantly more likely to mention that men are more numerically intelligent than women (47%) compared to those in the group discussion condition (10%), $t(38) = 2.88$, $p = .007$. There were also marginal differences in the same direction for beliefs that men have better spatial awareness than women $t(38) = 1.93$, $p = .06$. However, participants in the individual condition were also more likely to focus on positive in-group characteristics than when those participants who engaged in discussion. Participants in the individual condition were marginally more likely to report that women are more verbally intelligent than men $t(38) = 1.77$, $p = .09$, and women are more creative than men $t(38) = 1.95$, $p = .06$ (Table 1).

In contrast, participants in the group discussion condition were significantly more likely to report theories which justified or explained why the stereotype was accurate (100%) compared to the individual thought condition (79%), $t(38) = -2.61$, $p = .01$. Specifically, participants were more likely to theorize through discussion (86%) than through individual thought (47%) that the 'gender gap' is caused by physical differences. Evolution was given as an explanation by 43% of participants in the discussion condition, but this was mentioned by marginally fewer participants (16%) in the individual condition $t(38) = -1.90, p = .07$. Thus, discussion appeared to give the stereotype a firm theoretical basis, compared to individual thought. There were no other significant differences between conditions (Table 1).

For the question, 'Why is it not true that men are better than women at maths?' participants listed 4.14 ideas on average. There were 13 distinct ideas mentioned in total (for a full summary of the data, see Table 2). There was a significant difference between the conditions were on mentions of stereotype-disconfirming behaviour $t(34) = 3.93$, $p < .001$. Specifically, in the group discussion condition, 72% of women mentioned the behaviour that girls do better in exams/at school than boys, compared to only 17% in the individual condition. This suggests that the focus of discussions (relative to individual thoughts) was on positive in-group characteristics: a reverse of the finding in the affirm conditions where more positive in-group traits were written down after individual thought.

The frequency of theories which de-legitimized the purported gender difference also differed $t(34) = 3.81$, $p = .001$, with 89% reporting at least one theory in the discussion condition, compared to 33% in the individual condition. Examination of theories used to de-legitimize the gender stereotypes across the two challenge conditions suggested that group discussion appeared to focus participants' attention on the theory that any purported gender difference was due to prejudices in society $t(34) = 3.43$, $p = .002$ (83% compared to only 35% in the individual condition). Furthermore, 33% of participants in the group discussion conditions argued that environmental factors (not gender) determine numerical skill $t(34) = 2.19, p = .04$, compared to 6% in the individual thought condition. There were no significant differences between conditions on the remaining codes (Table 2). However, overall the significant differences across codes in the individual thought and discussion conditions provided support for Hypothesis (a): discussions resulted in qualitatively different content of deliberations relative to individual rumination.

Maths performance

In the initial model, maths scores were entered at level 1 and the dummy variables were entered at level 2. This initial model was significantly different to the null model,

$\chi^2(22) = 41.13$, $p < .01$, indicating significant between-condition differences overall. Analyses using the dummy variables indicated that there were significantly higher maths scores overall after the stereotype was challenged ($M = 2.42$, $SD = 1.05$), compared to when it was affirmed ($M = 1.79$, $SD = 1.10$), $\gamma = -0.62$, $p = .03$. There was no overall difference between group discussion ($M = 2.05$, $SD = 1.25$) and individual reflection ($M = 2.14$, $SD = 0.98$), $\gamma = -0.05$, $p = .86$. However, there was a significant two-way interaction, $\gamma = -1.50$, $p = .01$.

In order to explore this interaction, between-cell comparisons were conducted using further dummy and contrast variables. When participants challenged the stereotype through discussion ($M = 2.78$, $SD = 1.06$), they had significantly higher scores than when they affirmed the stereotype through discussion ($M = 1.40$, $SD = 1.05$), $\gamma = -1.38$, $p < .01$. Overall, participants who affirmed the stereotype through group discussion performed significantly worse than participants who affirmed the stereotype individually ($M = 2.21$, $SD = 1.03$), $\gamma = -0.79$, $p = .04$. Participants had marginally higher scores when they challenged the stereotype through discussion than alone ($M = 2.06$, $SD = 0.94$), $\gamma = -0.71$, $p = .07$. There was no significant difference between the challenge discussion and affirm individual conditions $\gamma = -0.58$, $p = .15$. This pattern of results in the interaction confirmed Hypothesis (b).

Stereotype consensus

A 2 (Topic: affirm stereotype vs. challenge stereotype) $\times$ 2 (Reflection: group discussion vs. individual thought) between-subjects ANOVA was conducted on the group-level consensus scores generated from the content analysis. Analyses revealed a main effect for reflection condition: there was greater consensus after discussion ($M = 0.91$, $SD = 0.10$) compared to individual thought ($M = 0.77$, $SD = 0.04$), $F(1,21) = 24.05$, $p < .01$, $\eta_p^2 = .53$, providing further support for Hypothesis (a). There was also significantly more consensus when participants affirmed the stereotype ($M = 0.87$, $SD = 0.10$) than challenged it ($M = 0.81$, $SD = 0.10$), $F(1,21) = 4.33$, $p = .05$, $\eta_p^2 = .17$. The two-way interaction was not significant $F(1,21) = 1.58$, $p = .22$, $\eta_p^2 = .07$.

Simple cell analyses of the consensus score revealed no difference when affirming or challenging the stereotype through individual thought $F(1,21) = 0.33$, $p = .57$, $\eta_p^2 = .02$. However, there was significantly more consensus when affirming the stereotype through discussion than challenging it though discussion $F(1,21) = 5.78$, $p = .03$, $\eta_p^2 = .22$. There was also significantly more consensus about the stereotype after affirming it through discussion ($M = 0.95$, $SD = 0.05$) compared to individual thought ($M = 0.78$, $SD = 0.05$), $F(1,21) = 19.71$, $p < .01$, $\eta_p^2 = .48$. There was also a greater degree of consensus after group discussion which contested the stereotype ($M = 0.86$, $SD = 0.12$) compared to individual reflection ($M = 0.76$, $SD = 0.02$), $F(1,21) = 6.41$, $p = .02$, $\eta_p^2 = .23$.

Mediation

In order to establish whether group consensus on the stereotype mediated the effect of discussion versus individual thought on maths performance when affirming or contesting the stereotype, we conducted analyses of covariance using the group consensus scores and maths performance at the group level. When contesting the stereotype, consensus was not significant covariate $F(1, 21) = 0.07$, $p = .81$, $\eta_p^2 = .02$. However, when affirming the stereotype, consensus significantly co-varied with maths score $F(1, 21) = 6.69$,

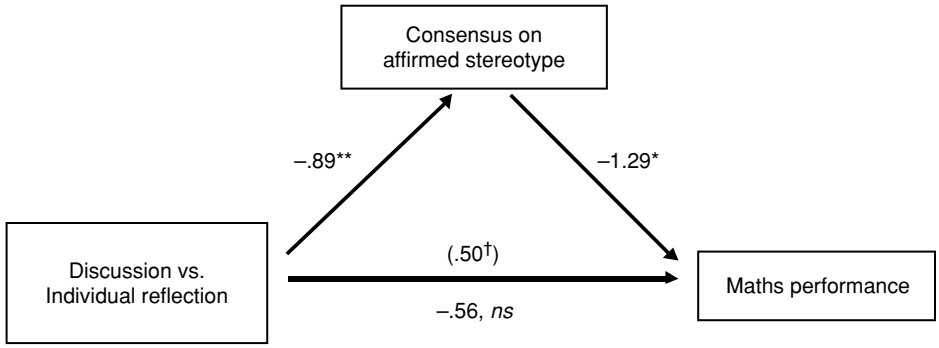

Figure 1. Beta coefficients for mediation model in affirm conditions, Study 1 ($^{\dagger}.07$; $^{*}p < .05$; $^{**}p < .01$).

$p = .03$, $\eta_p^2 = .50$. On entering this mediator into the model, the difference between participants who affirmed the stereotype through discussion or individual reflection became non-significant $F(1, 21) = 1.31$, $p = .29$, $\eta_p^2 = .16$. A Sobel test indicated that this was a significant reduction, $z = 2.62$, $p < .01$, providing evidence that consensus on the affirmed stereotype mediated the effects of discussion relative to individual rumination on maths score, providing partial support for Hypothesis (c) (Figure 1).

Discussion

The purpose of this study was to establish the effects of group discussion (vs. individual thought) about relevant gender stereotypes on performance on a maths test. We predicted an interaction between type of reflection (discussion vs. individual thought) and topic of reflection (affirming or challenging the stereotype). Specifically, we hypothesized that participants who affirmed the stereotype would have lower maths scores than participants who challenged the stereotype, but this difference would be achieved through group discussion rather than individual cognition. In support of this theory, there was a difference between the maths performance of participants who affirmed and challenged the stereotype, but this was only significant among participants who had discussed the topic, not among participants who reflected on it individually. Furthermore, there was a marginal lift in performance after discussions that challenged the stereotype relative to the individual challenge condition, tentatively suggesting that group discussion provided a cognitive boost that individual cognition alone could not provide.

We also predicted that group discussion would have this effect because it would provide participants with the opportunity to exchange and validate views as to whether or not the stereotype is accurate (Festinger, 1950, 1954). In line with this hypothesis, the content of the stereotype in the affirmative discussions supported the legitimacy of the stereotype. Participants' explanations as to why women were worse than men at maths were saturated with stereotyped gender statements (cf. McGarty, Yzerbyt, & Spears, 2002), and these were qualitatively different and more consensual after group discussion compared to individual thought. Participants who affirmed the issue through

discussion appeared more likely to provide (essentialising) physiological theories that would explain the negative performance of their group. In this way, group discussion accentuated and entrenched participants' views of the stereotype which suggested their performance should be lower. Therefore, a consensual confirmatory discussion appeared to restrict participants to the mould of the stereotype. Importantly, this phenomenon mediated participants' lower performance in this condition, providing evidence that this was indeed the driving process.

In the condition in which the stereotype was challenged, group discussion seemed to focus participants' attention on positive group characteristics and de-legitimize constraining factors which may determine mathematical skills, like physical differences. Participants in this condition had more consensual views on how to challenge the societal stereotypes, although consensus did not mediate the difference between conditions when the stereotypes were challenged. Instead, we speculated that invalidating or de-legitimizing the stereotype (as was evident in the content in these conditions) may have increased participants' confidence in their mathematical ability. After all, previous research has suggested that group discussion can be a source of action confidence (Baron *et al.*, 1996; Luus & Wells, 1994; Petty *et al.*, 2002), and Else-Quest *et al.*'s (2010) meta-analysis showed that males had greater mathematical confidence than females, alongside their higher test scores. Therefore, we speculated that the process of challenging the stereotype together (due to its power to invalidate the stereotype) could give women more confidence in their mathematical ability than those who affirm the stereotype, and this may lead to better performance.

Although the effects of Study 1 were encouraging, the evaluation of the direction and strength of the effects of group discussion were limited by (a) the absence of a control condition in which participants did not explicitly reflect on stereotypes, and therefore a baseline against which to evaluate effects of group discussion and (b) the absence of conditions in which male participants who underwent the same procedure. The question also remained as to whether contesting the stereotype in groups worked to increase performance (relative to the affirming discussions) because of a change in confidence in mathematical ability. Study 2 was designed to investigate these issues.

STUDY 2

In the second study, we compared the effects of small group discussion on the maths performance of both women and men, and included a baseline condition in which participants did not explicitly reflect on the stereotype. We hypothesized that (a) gender and discussion topic would interact: women who affirmed the stereotype during a group discussion would show decreased maths performance (relative to the challenge and baseline conditions). In contrast, men who affirmed the stereotype during group discussion would show improved performance (stereotype lift) relative to baseline scores. When men and women were given the opportunity to challenge the stereotype through discussion, we predicted that there would be no difference on maths performance between the gender groups. A final purpose of this study was to investigate whether challenging a stereotype through discussion (compared to individual thought) can work to increase individuals' confidence in their maths ability, improving performance. Accordingly, we hypothesized (c) that confidence in ability would mediate the effect of experimental condition on maths performance.

Method

Participants and design

Participants were 255 undergraduate volunteers from the same university as before (134 male, 121 female) and the mean age was 19.68 years ($SD = 1.41$, range = 13). All participants were unpaid and naive as to the purposes of the study. Maths ability was controlled by only including participants who did not study maths at a degree level. Participants were randomly assigned to one of the conditions of the 3 (Discussion topic: affirm stereotype vs. challenge stereotype vs. no discussion baseline) $\times$ 2 (Gender: male vs. female) between-subjects factorial design. Participants were randomly allocated to same-gender three-person discussion groups ($N = 80$) by assigning each individual a group number.

Materials and procedure

The procedure in the experimental conditions was identical to Study 1, except that participants did not write down their responses, and therefore there were no qualitative data to analyse. In the baseline condition, participants were not given any feedback on the stereotype or told that the test was diagnostic, were merely asked to do their best on the test, and had no opportunity to discuss the study at any point.

Dependent measures

The primary dependent measure was the total score on the maths test, which was expanded from Study 1 to contain 11 items, to increase the sensitivity of the test. The additional items were, '$-9 + (-7) = ?$'; '$3/5 - 1/4 = ?$'; 'A student got 90 out of 270 in a maths test. What percentage is this?' and 'Simplify $18^6 \div 18^2$'. Correct answers were totalled to give a score out of 11. In addition to the maths test, participants responded to a single item, 'I am confident in my mathematical ability'. Participants responded to the item on a Likert-type scale, (1 = 'strongly disagree', to 7 = 'strongly agree') to what extent they agreed or disagreed with the statement.

Analytic strategy

HLM analyses were performed on the data in the 2 (Discussion topic: affirm stereotype vs. challenge stereotype) $\times$ 2 (Gender: male vs. female) design, in HLM 6.03 (Raudenbush & Bryk, 2002). Main effects and the two-way interaction were tested using dummy variables (Judd *et al.*, 1995; Rosenthal & Rosnow, 1985; Wilkinson *et al.*, 1999). First, the affirm stereotype conditions [0] were compared to the challenge stereotype conditions [+1], for men [0] and women [+1]. Second, to interpret the main effect of discussion topic on maths scores, we performed a series of contrast analyses on additional models, which included the baseline control conditions (cf. Poortvliet, Janssen, Van Yperen, & Van de Vliert, 2007).

Results

Maths performance

In the initial model, maths scores were entered at level 1 and the dummy variables were entered at level 2. This initial model was significantly different to the null model, $\chi^2\ (53) = 72.54$, $p = .04$, indicating significant between-condition differences overall.

There was a significant gender main effect, $\gamma = -3.42$, $p = .02$ and a marginal main effect for topic of discussion $\gamma = -2.86$, $p = .06$. Please refer to Table 4 for means. These effects were qualified however, by a significant two-way interaction between the topic of discussion and gender $\gamma = 2.05$, $p = .03$. Women ($M = 6.65$, $SD = 2.62$) scored more than men ($M = 6.04$, $SD = 3.01$) when participants of both genders challenged the stereotype, and men ($M = 6.85$, $SD = 2.69$) scored more highly than women ($M = 5.42$, $SD = 2.82$) when participants of both genders affirmed the stereotype.

Table 4. Means and standard deviations of dependent variables, Study 2

	Male			Female		
Discussion condition	Affirmed stereotype	Challenged stereotype	Baseline control	Affirmed stereotype	Challenged stereotype	Baseline control
Maths test score/11						
M	6.85_b	6.04_{ab}	5.39_a	5.42_a	6.65_b	6.36_b
SD	2.69	3.01	2.66	2.82	2.62	2.77
Confidence in mathematical ability/7						
M	4.32_b	4.08_b	4.26_b	3.09_a	4.00_b	3.82_b
SD	1.75	1.98	1.88	1.91	1.74	2.24

Note. Means in rows with different subscripts differ at $p < .05$.

In the baseline control condition, there was no difference between female and male participants on maths performance $\gamma = -1.38$, $p = .11$. However, after discussions affirming the stereotype, women performed significantly more poorly than men, $\gamma = -1.25$, $p = .04$. After a discussion which challenged the stereotype, women and men performed equally well, $\gamma = 0.81$, $p = .26$.

Examining the between-condition effects for women, a group discussion to affirm the stereotype produced significantly lower scores on the maths test ($M = 5.42$, $SD = 2.82$) compared with the baseline scores ($M = 6.36$, $SD = 2.77$), $\gamma = -1.34$, $p = .05$. When female participants challenged the stereotype, their scores were no different ($M = 6.65$, $SD = 2.62$) than scores of female participants in the baseline control condition ($M = 6.36$, $SD = 2.77$), $\gamma = -0.03$, $p = .97$. Finally, women who challenged the stereotype had higher maths test scores ($M = 6.65$, $SD = 2.62$) than women who affirmed the stereotype ($M = 5.42$, $SD = 2.82$) $\gamma = -1.25$, $p = .04$.

For men, a discussion that affirmed the stereotype led to higher scores on the maths test ($M = 6.85$, $SD = 2.69$) than the baseline ($M = 5.39$, $SD = 2.66$), $\gamma = 1.46$, $p = .03$ providing evidence for stereotype lift. When group discussion challenged the stereotype, maths test performance ($M = 6.04$, $SD = 3.01$) was not significantly different from the baseline ($M = 5.39$, $SD = 2.66$), $\gamma = 0.60$, $p = .44$. It is notable, however, that even though the discussion was meant to challenge the stereotype, performance of men was somewhat higher than baseline. There was no significant difference between maths scores for men, $\gamma = 0.81$, $p = .26$, when they affirmed the stereotype ($M = 6.85$, $SD = 2.69$) or challenged it ($M = 6.04$, $SD = 3.01$), although scores in the affirm condition were somewhat elevated.

Confidence in mathematical ability

Responses to the item, 'I am confident in my mathematical ability' were entered into the model at level 1 and the dummy variables were entered at level 2. The gender main

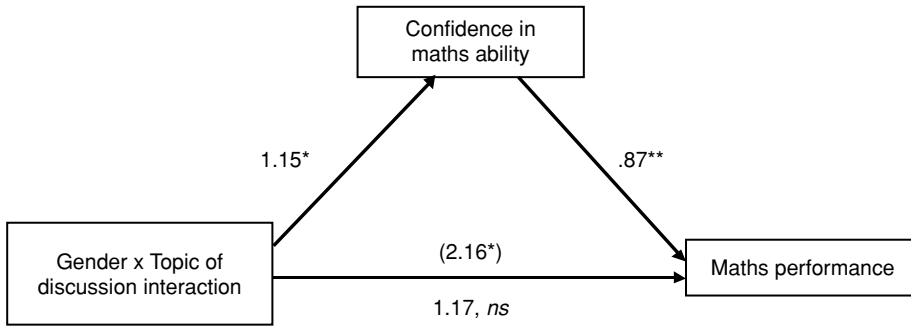

Figure 2. Gamma coefficients for mediation model, Study 2 (*$p < .05$; **$p < .01$).

effect was not significant, $\gamma = 0.02$, $p = .96$, but there was a significant main effect for topic of discussion $\gamma = -0.94$, $p < .01$, with more confidence overall when participants contested ($M = 4.04$, $SD = 1.86$) than affirmed ($M = 3.67$, $SD = 1.93$) the stereotype. These effects were qualified as before, by a significant two-way interaction between the topic of discussion and gender $\gamma = 1.15$, $p < .05$. Contrasts showed that men ($M = 4.32$, $SD = 1.75$) were more confident in their abilities than women ($M = 3.09$, $SD = 1.91$) when participants of both genders affirmed the stereotype $\gamma = -1.26$, $p < .01$, and scores were no different for men ($M = 4.08$, $SD = 1.98$) and women ($M = 4.00$, $SD = 1.74$) when participants of both genders contested the stereotype, $\gamma = -0.04$, $p = .93$.

Examining the between-condition effects for women, a group discussion to affirm the stereotype produced marginally lower confidence in maths ability compared with the baseline scores ($M = 3.82$, $SD = 2.24$), $\gamma = -0.94$, $p = .07$. Women who challenged the stereotype had significantly more confidence than women who affirmed the stereotype $\gamma = -0.93$, $p = .03$. When female participants challenged the stereotype, their confidence was equal to that of female participants in the baseline control condition, $\gamma = 0.08$, $p = .90$.

For men, a discussion that affirmed $\gamma = 0.13$, $p = .77$ or challenged $\gamma = -0.12$, $p = .84$ the stereotype led to no change in confidence relative to the baseline ($M = 4.26$, $SD = 2.66$). There was also no significant difference in confidence whether men affirmed or challenged the stereotype $\gamma = 0.21$, $p = .63$.

Mediation[3]

Next, confidence in ability was entered into the model as a grand centred predictor at level 1, predicting maths scores in the experimental conditions. The slope was significant $\gamma = 0.88$, $p < .01$. As confidence was a significant predictor of maths scores and also differed across conditions, it was entered as a level 1 predictor, with maths score as the outcome variable. The dummy variables for the main effects and interaction were entered separately at level 2. The slope for confidence was significant $\gamma = 0.87$,

[3] *We used the multi-level first-order Taylor series approximation to estimate the standard error of the mediated effect as is recommended for multi-level mediation with a level 2 predictor, level 1 mediator, and level 1 outcome (see Krull & MacKinnon, 1999).*

$p < .01$. On entering this mediator into the model, the two-way interaction became non-significant $\gamma = 1.17$, $p = .15$, Sobel $z = 1.97$, $p < .05$; suggesting confidence in ability mediated the effects of condition on maths performance (Figure 2).

Discussion

The aim of the second study was to investigate the effects of group discussion on both women's and men's maths performance, and compare these to baseline scores. First, we predicted an interaction between gender and discussion topic. As hypothesized, results showed that women performed significantly worse than men and than their baseline score when they had affirmed the stereotype during discussion. Male participants' maths scores, however, were lifted when they affirmed the stereotype through discussion.

Second, it was predicted that when men and women were given the opportunity to challenge the stereotype through discussion, there should be no difference on maths performance between the gender groups. In line with this hypothesis, results showed that women did no worse than men after a group discussion which challenged the stereotype. In fact, the effect was slightly reversed when women had challenged the stereotype in their group. Therefore, whilst discussion reinforced the stereotype when it was affirmed, it also appeared to be able to equip women to resist its negative effects.

In light of the finding of the previous study that consensus did not drive the effects on performance in the challenge conditions, we investigated an alternative mediator in Study 2: confidence in ability. Our results showed that confidence mediated the effects of experimental condition on performance. This is in line with previous research which points to the role of corroboration by other group members in increasing validation and confidence (Baron *et al.*, 1996; Luus & Wells, 1994; Petty *et al.*, 2002), and supports our explanation of why discussion should impact in this way on behaviour: because of the critical role of the opinion of others in socially shared phenomena such as stereotyping and their effects.

GENERAL DISCUSSION

We conducted two studies in order to provide insight into the effect of group discussion about a stereotype on the performance of targeted individuals. In order to maximize the pertinence and timeliness of this research, we situated it in the context of a stereotype that has been the subject of much recent public debate (Dar-Nimrod & Heine, 2006) and psychological research (Marx & Roman, 2002; Nosek *et al.*, 2002; Schmader, 2002; Spencer *et al.*, 1999): gender and mathematical ability.

In the first study, we investigated the unique consequences of discussion about the stereotypes (compared to individual thought) on behaviour. We found that as expected, discussion altered performance relative to individual thought, and the degree to which the content of the discussions that validated the stereotype were consensual mediated the impact of these discussions on performance. In other words, when other group members agreed that the stereotype was valid and had a legitimate basis, performance was poor compared to when participants thought about the stereotype on their own. This finding supports previous theorizing that other group members provide a resource through which we can validate and confirm social reality (Correll & Park, 2005; Festinger, 1954; Macrae, Bodenhausen, Milne, & Jetten, 1994), and this validated knowledge can

guide behaviour (Abrams & Hogg, 1990; Haslam *et al.*, 1999; Reicher *et al.*, 1995; Sechrist & Stangor, 2001; Turner, 1985).

Fundamentally, by demonstrating that the power of stereotypes to affect behaviour is enhanced by their being validated by others and by manipulating this process, we have highlighted that a powerful aspect stereotyping lies in the fact that they are socially shared and socially legitimized, and therefore can be socially de-legitimized. Furthermore, the results suggest that strategies of legitimizing and de-legitimizing stereotypes are more effective when conducted with others than through isolated individual cognitions. Group discussion enabled group members to exchange validating ideas about group characteristics and therefore gain valuable, socially shared information about group attitudes and behaviour. To our knowledge, this is the first intervention in the stereotyping literature that focuses on changing perceptions of the validity of stereotypes through interaction between members of the stereotyped group.

As supported by the evidence of stereotype content, the shackles of stigma appeared to only apply when participants agreed upon the validity of the stereotype together. When female participants were invited to challenge the negative stereotype of women, they appeared to use the brief group discussion to come up with explanations (or theories) which invalidated the stereotype or made it more malleable. Indeed, there was a promising marginal lift in maths performance for women who challenged the stereotype collectively relative to alone. We suggested that this could be because hearing group members voice theories as to why the stereotype was invalid have increased participants' confidence in their abilities relative to those participants who heard the stereotype being validated, and tested this explanation in the subsequent study.

In the second study, we compared the impact of group discussion on the performance of both the positively and negatively stereotyped targets, and investigated whether a change in confidence could explain the performance difference in the affirm and challenge conditions. Feedback from peers has been suggested to be a source of action confidence (Baron *et al.*, 1996; Luus & Wells, 1994; Petty *et al.*, 2002), and the results of Study 2 confirmed this, with greater confidence and better maths performance when women challenged the stereotype. The findings for male participants were also reminiscent of the recently described phenomenon of stereotype 'lift' (Walton & Cohen, 2003), whereby affirming and this validating the stereotyping of the out-group (women) was utilized to enhance in-group (men's) performance.

Theoretical implications

Our results suggest that if a group member perceives that the stereotype of their group is validated and agreed upon by other group members, their performance could be implicitly guided by the stereotype. This finding supports previous research on stereotyping that has noted the importance of viewing stereotypes as socially shared representations (Kashima, 2000; Lyons & Kashima, 2003). This relocates the process of stereotyping to within the intragroup dynamic (Carnaghi & Yzerbyt, 2007) and provides this research with phenomenological resonance and ecological validity. As such, the findings help support the meta-theoretical centrality of intragroup dynamics and communication as key predictors of the impact of stereotypes on behaviour. From a wider perspective, these findings help to integrate work in the stereotype threat domain with small group research on stereotyping and stereotype consensus, showing that negative in-group stereotypes can increase in consensus during discussion and demonstrating that these stereotypes can affect behaviour.

Practical implications

The results presented here are a promising (albeit preliminary) demonstration that social and behavioural change could be affected in small groups by discussing a stereotype. In other words, stereotype threat may be less likely to occur if targeted individuals are able to challenge the stereotype with others before taking a relevant performance test. This is especially pertinent in the light of recent high school exam results in the UK, and of the cross-national meta-analysis by Else-Quest *et al.* (2010), both of which show that boys are outperforming girls in mathematics exams. Our results show that brief, small group discussions with fellow targets of prejudice can be effective in helping them (in the short term at least), to overcome the impact of negative self-stereotypes in the experimental context. Social psychologically, it is interesting to highlight the implications of hearing shared opinions about the validity of stereotypes, and by doing so re-focusing the study of stereotypes within the intragroup dynamic (Carnaghi & Yzerbyt, 2007).

Future directions

The next stage in this research programme would be to test whether repeated discussion on this topic over time could instigate chronic stereotype change and an associated stable performance change across contexts (cf. Paluck, 2009). It is important to know whether micro-conversations that are repeated and reinforced over time are powerful enough to create stereotype change that can endure through exposure to occasional contradictory conversations (the effect of which we demonstrate in the present research). Current knowledge concerning implicit stereotyping might warn us against necessarily drawing optimistic conclusions from evidence of short-term effects of participating in discussion aimed at negating a particular stereotype (e.g., rebound effects; Macrae *et al.*, 1994). The longer-term project described above may help us to understand whether dialogue results in or inoculates against such effects. Before such a logistically complex long-term project could take place however, it was first necessary to demonstrate the power of small group conversations in the short term, as we have here. The future research described above would help to bridge the divide between micro-experimental contexts and snapshot interventions, and real-time social change (see Condor, 1996). It is therefore, imperative that future research investigates further the process through which small group interaction achieves this change in the long term. The reported studies are a first step, and future work will need to refine such methods and extend them beyond the laboratory over time to test their applicability to educational policy and practice.

Conclusion

Research has robustly demonstrated the tendency for activated stereotypes to affect the behaviour of targeted individuals on relevant dimensions (e.g., Aronson *et al.*, 1999; Steele, 1997; Steele & Aronson, 1995). These stereotypes are socially shared representations of social groups (Kashima, 2000; Lyons & Kashima, 2003), and as such are often discussed in society (e.g., Karasawa *et al.*, 2007; Ruscher, 2001; van Dijk, 1987). However, to date, no research has examined the impact of discussing stereotypes on their ability to affect performance. In this research, we systematically examined the affect of small group interaction on the impact of stereotyping on behaviour. The results of the two experiments presented here offer support for the notion that nuanced small

group interaction interventions should be studied in the future in an attempt to reduce the impact of stereotypes on performance. Our experiments found that participants who discussed reasons why the stereotype was invalid, compared to those who affirmed its validity in groups or did not discuss the stereotype, completed more maths questions correctly and had greater confidence in their abilities. These findings support the notion that as socially shared representations, stereotypes derive their power from being agreed upon and validated by others. Any perception that other people do not support the stereotype's validity can work to undermine its power to influence the performance of individuals.

References

Abrams, D., & Hogg, M. A. (1990). Social identification, self-categorization and social influence. *European Review of Social Psychology*, *1*, 195–228. doi:10.1080/14792779108401862

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, *38*, 113–125. doi:10.1006/jesp.2001.1491

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White men can't do maths: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, *35*, 29–46. doi:10.1006/jesp.1998.1371

Baron, R. S., Hoppe, S. I., Kao, C. F., Brunsman, B., Linneweh, B., & Rogers, D. (1996). Social corroboration and opinion extremity. *Journal of Experimental Social Psychology*, *32*, 537–560. doi:10.1006/jesp.1996.0024

Carnaghi, A., & Yzerbyt, V. Y. (2007). Subtyping and social consensus: The role of the audience in the maintenance of stereotypic beliefs. *European Journal of Social Psychology*, *37*, 902–922. doi:10.1002/ejsp.402

Chong, D., & Druckman, J. N. (2007). Framing public opinion in competitive democracies. *American Political Science Review*, *101*, 637–655. doi:10.1017/S0003055407070554

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J. F. Leny & W. Kintsch (Eds.), *Language and comprehension* (pp. 287–299). Amsterdam: North-Holland.

Cockcroft, W. H. (1982). *Mathematics counts: Report of the committee of inquiry into the teaching of mathematics in schools*. London: HMSO.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46. doi:10.1177/001316446002000104

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioural sciences* (3rd ed.). Hillsdale, NJ: Erlbaum.

Condor, S. (1996). Social identity and time. In W. P. Robinson (Ed.), *Social groups and identities: Developing the legacy of Henri Tajfel* (pp. 285–315). Oxford: Butterworth-Heinemann.

Correll, J., & Park, B. (2005). A model of the in-group as a social resource. *Personality and Social Psychology Review*, *9*(4), 341–359. doi:10.1207/s15327957pspr0904_4

Craemer, T. (2008). Nonconscious feelings of closeness toward African Americans and support for pro-Black policies. *Political Psychology*, *29*, 407–435. doi:10.1111/j.1467-9221.2008.00637.x

Croizet, J.-C., Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, *24*, 588–594. doi:10.1177/0146167298246003

Curtis, P. (2009). *GCSE results: Boys bag top grades in maths*. Retrieved from http://www.guardian.co.uk/education/2009/aug/27/gcse-results-2009-maths-exams

Dar-Nimrod, I., & Heine, S. J. (2006). Exposure to scientific theories affects women's math performance. *Science*, *314*, 435. doi:10.1126/science.1131100

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and

professionally. *Personality and Social Psychology Bulletin*, *28*, 1615–1628. doi:10.1177/014616702237644

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*, 103–127. doi:10.1037/a0018053

Festinger, L. (1950). Informal social communication. *Psychological Review*, *57*, 271–282. doi:10.1037/h0056932

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–140. doi:10.1177/001872675400700202

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, *24*, 645–662. doi:10.1016/j.appdev.2003.09.002

Hardin, C. D., & Higgins, E. T. (1996). Shared reality: How social verification makes the subjective objective. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition: The interpersonal context* (Vol 3, pp. 28–84). New York: Guilford.

Haslam, S. A. (1997). Stereotyping and social influence: Foundations of stereotype consensus. In R. Spears P. J. Oakes, N. Ellemers, & S. A. Haslam (Eds.), *The social psychology of stereotyping and group life* (pp. 119–143). Oxford: Blackwell.

Haslam, S. A., Oakes, P. J., Reynolds, K. J., & Turner, J. C. (1999). Social identity salience and the emergence of stereotype consensus. *Personality and Social Psychology Bulletin*, *25*(7), 809–818. doi:10.1177/0146167299025007004

Haslam, S. A., Salvatore, J., Kessler, T., & Reicher, S. D. (2008). The social psychology of success. *Scientific American Mind*, *19*(2), 24–31. doi:10.1038/scientificamericanmind0408-24

Haslam, S. A., Turner, J. C., Oakes, P. J., McGarty, C., & Reynolds, K. J. (1998). The group as a basis for emergent stereotype consensus. *European Review of Social Psychology*, *28*, 755–776. doi:10.1002/(SICI)1099-0992(199809/10)28:5<755::AID-EJSP891>3.0.CO;2-Z

Haslam, S. A., Turner, J. C., Oakes, P. J., Reynolds, K. J., Eggins, R. A., Nolan, M., & Tweedie, J. (1998). When do stereotypes become really consensual? Investigating the group-based dynamics of the consensualisation process. *European Journal of Social Psychology*, *28*, 755–776. doi:10.1002/(SICI)1099-0992(199809/10)28:5<755::AID-EJSP891>3.0.CO;2-Z

Haworth, C. M. A., Dale, P., & Plomin, R. (2009). Sex differences and science: The etiology of science excellence. *Journal of Child Psychology and Psychiatry*, *50*, 1113–1120. doi:10.1111/j.1469-7610.2009.02087.x

Higgins, E. T. (1992). Achieving 'shared reality' in the communication game: A social action that creates meaning. *Journal of Language and Social Psychology*, *11*, 107–131. doi:10.1177/0261927X92113001

Judd, C. M., McClelland, G. H., & Culhane, S. C. (1995). Data analysis: Continuing issues in the everyday analysis of psychological data. *Annual Review of Psychology*, *46*, 433–465. doi:10.1146/annurev.ps.46.020195.002245

Karasawa, M., Asai, N., & Tanabe, Y. (2007). Stereotypes as shared beliefs: Effects of group identity on dyadic conversations. *Group Processes and Intergroup Relations*, *10*, 515–532. doi:10.1177/1368430207081538

Kashima, Y. (2000). Maintaining cultural stereotypes in the serial reproduction of narratives. *Personality and Social Psychology Bulletin*, *26*, 594–604. doi:10.1177/0146167200267007

Krull, T. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, *23*, 418–444.

Lewin, K. (1953). Studies in group decision. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory* (pp. 287–301). New York: Row, Peterson.

Luus, E. C. A., & Wells, G. L. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. *Journal of Applied Psychology*, *79*, 714–723. doi:10.1037/0021-9010.79.5.714

Lyons, A., & Kashima, Y. (2003). How are stereotypes maintained through communication? The influence of stereotype sharedness. *Journal of Personality and Social Psychology*, *85*, 989–1005. doi:10.1037/0022-3514.85.6.989

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology*, *67*, 808–817. doi:10.1037/0022-3514.67.5.808

Mallett, R. K., Huntsinger, J. R., Sinclair, S., & Swim, J. K. (2008). Seeing through their eyes: When majority group members take collective action on behalf of an outgroup. *Group Processes and Intergroup Relations*, *11*, 451–470. doi:10.1177/1368430208095400

Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology*, *42*, 236–243. doi:10.1016/j.jesp.2005.04.010

Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin*, *28*, 1183–1193. doi:10.1177/01461672022812004

McGarty, C., Turner, J. C., Oakes, P. J., & Haslam, S. A. (1993). The creation of uncertainty in the influence process: The roles of stimulus information and disagreement with similar others. *European Journal of Social Psychology*, *23*, 17–38. doi:10.1002/ejsp.2420230103

McGarty, C., Yzerbyt, V. Y., & Spears, R. (2002). *Stereotypes as explanations: The formation of meaningful beliefs about social groups*. Cambridge: Cambridge University Press.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). London: Sage.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology*, *83*, 44–59. doi:10.1037/0022-3514.83.1.44

Ouwerkerk, J. W., de Gilder, D., & de Vries, N. K. (2000). When the going gets tough, the tough get going: Social identification and individual effort in intergroup competition. *Personality and Social Psychology Bulletin*, *26*, 1550–1559. doi:10.1177/01461672002612009

Paluck, E. L. (2009). What's in a norm? Sources and processes of norm change. *Journal of Personality and Social Psychology*, *96*(3), 594–600. doi:10.1037/a0014688

Petty, R. E., Briñol, P., & Tormala, Z. L. (2002). Thought confidence as a determinant of persuasion: The self-validation hypothesis. *Journal of Personality and Social Psychology*, *82*, 722–741. doi:10.1037/0022-3514.82.5.722

Poortvliet, P. M., Janssen, O., Van Yperen, N. W., & Van de Vliert, E. (2007). Achievement goals and interpersonal behaviour: How mastery and performance goals shape information exchange. *Personality and Social Psychology Bulletin*, *33*(10), 1435–1447.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Reicher, S. D., Spears, R., & Postmes, T. (1995). More on deindividuation, power relations between groups and the expression of social identity: Three studies on the effects of visibility to the in-group. *European Review of Social Psychology*, *6*, 161–168. doi:10.1080/14792779443000049

Rosenthal, H. E. S., & Crisp, R. J. (2006). Reducing stereotype threat by blurring intergroup boundaries. *Personality and Social Psychology Bulletin*, *32*, 501–511. doi:10.1177/0146167205281009

Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analyses: Focused comparisons in the analysis of variance*. Cambridge: Cambridge University Press.

Ruscher, J. B. (2001). *Prejudiced communication: A social psychological perspective*. New York: Guilford.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, *38*, 194–201. doi:10.1006/jesp.2001.1500

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*(2), 336–356. doi:10.1037/0033-295X.115.2.336

Sechrist, G. B., & Stangor, C. (2001). Perceived consensus influences intergroup behavior and stereotype accessibility. *Journal of Personality and Social Psychology*, *2*, 645–654. doi:10.1037/0022-3514.80.4.645

Semin, G. R. (2008). Stereotypes in the wild. In Y., Kashima, K. Fiedler, & P. Freytag (Eds.), *Stereotype dynamics: Language-based approaches to the formation, maintenance and transformation of stereotypes* (pp. 11–28). London: Erlbaum.

Shapiro, J. R., & Neuberg, S. L. (2007). From stereotype threat to stereotype threats: Implications of a multi-threat framework for causes, moderators, mediators, consequences, and interventions. *Personality and Social Psychology Review*, *11*, 107–130. doi:10.1177/1088868306294790

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioural impulses. *Psychological Review*, *115*, 314–335. doi:10.1037/0033-295X.115.2.314

Smith, J. L., & White, P. H. (2002). An examination of implicitly activated, explicitly activated, and nullified stereotypes on mathematical performance: It's not just a woman's issue. *Sex Roles*, *47*, 179–191. doi:10.1023/A:1021051223441

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's maths performance. *Journal of Experimental Social Psychology*, *35*, 4–28. doi:10.1006/jesp.1998.1373

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *57*(6), 613–629. doi:10.1037/0003-066X.52.6.613

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *64*, 885–896. doi:10.1037/0022-3514.64.6.885

Stott, C., & Drury, J. (2004). The importance of social structure and social interaction in stereotype consensus and content. *European Journal of Social Psychology*, *34*, 11–24. doi:10.1002/ejsp.183

Subasic, E., Reynolds, K. J., & Turner, J. C. (2009). The political solidarity model of social change: Dynamics of self-categorization in intergroup power relations. *Personality and Social Psychology Review*, *12*, 330–352. doi:10.1177/1088868308323223

Thompson, M. S., Judd, C. M., & Park, B. (2000). The consequences of communicating social stereotypes. *Journal of Experimental Social Psychology*, *36*, 567–599. doi:10.1006/jesp.1999.1419

Turner, J. C. (1985). Social categorization and the self concept: A social cognitive theory of group behaviour. In E. J. Lawler (Ed.), *Advances in group processes: Theory and research* (Vol. 2, pp. 77–271). Greenwich, CT: JAL Press.

Turner, R. H., & Killian, L. M. (1972). *Collective behaviour*. Englewood Cliffs, NJ: Prentice-Hall.

van Dijk, T. (1987). *Communicating racism: Ethnic prejudice in thought and talk*. London: Sage.

van Zomeren, M., Postmes, T., & Spears, R. (2008). Toward an integrative social identity model of collective action: A quantitative research synthesis of three socio-psychological perspectives. *Psychological Bulletin*, *134*, 504–535. doi:10.1037/0033-2909.134.4.504

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, *39*, 456–467. doi:10.1016/S0022-1031(03)00019-2

Wilkinson, L., & The APA Task Force on Statistical Inference (1999). Statistical methods in psychology journals. *American Psychologist*, *54*, 594–604.

Received 30 March 2009; revised version received 8 February 2010