Article accepted 16th November 2020 in *Sport, Exercise and Performance Psychology*

Please cite this article as:

Rees, T. & Salvatore, J. (in press). Questioning stereotypes disrupts the effects of stereotype threat. *Sport, Exercise and Performance Psychology*. doi:10.1037/spy0000247.

Questioning Stereotypes Disrupts the Effects of Stereotype Threat

Tim Rees[1] and Jessica Salvatore[2]

[1]Department of Rehabilitation and Sport Sciences, Bournemouth University

[2]Department of Psychology, Sweet Briar College

Author Note

Tim Rees https://orcid.org/0000-0001-5498-0145

Jessica Salvatore https://orcid.org/0000-0003-1223-7008

We thank Marcus Pritchard, Benjamin Wedge, Amy Sambell, Stuart Tame, Jack Sanders, Jake Nicholls, and Lucas Sohl, for their help in data collection. We thank Thomas Morton and Pete Coffee for feedback on a prior version of this manuscript.

The datasets used in this article may be downloaded via the Open Science Foundation at https://osf.io/c8qyd/

Correspondence concerning this article should be addressed to Tim Rees, Department of Rehabilitation and Sport Sciences, Faculty of Health and Social Sciences, Bournemouth University, Talbot Campus, Fern Barrow, Poole, BH12 5BB, United Kingdom. Email: trees@bournemouth.ac.uk

Abstract

Given that mentally activating negative stereotypes about our social groups impairs our performance, can questioning the stereotype effectively disrupt this phenomenon? We experimentally tested an intervention of this type in several samples of athletes. Performance was consistently much improved, both in statistical and in practical terms, when participants were encouraged to question the stereotype than when they were merely reminded of the stereotype. These effects held regardless of the content of the stereotype and the group targeted. Process evidence suggests that questioning the stereotype buffers performance primarily via affect: the intervention stops athletes from worrying. Taken together, these findings indicate that adopting a simple yet powerful questioning stance protects targets of stereotypes against the performance impairments that they would otherwise typically experience.

Keywords: stereotype threat, athletic performance, intervention, questioning, challenge

Questioning Stereotypes Disrupts the Effects of Stereotype Threat

A ubiquitous feature of complex human social environments is that certain groups are stereotyped as inferior and others as superior. Activating individuals' awareness of this social fact impairs the performance of negatively stereotyped groups (Steele et al., 2002) and improves the performance of positively stereotyped groups (Walton & Cohen, 2003).

Sport is a relevant place to study how stereotypes are maintained through expectancy confirmation (see, e.g., Stone et al., 1997). More recently, attention has turned from perceivers to targets of stereotypes, and the fact that athletes "perform" is now central to this research program, rather than incidental. That stereotypes impair sporting performance in particular—alongside intellectual performance, where the phenomenon was first observed—is now well documented (for reviews, see Beilock & McConnell, 2004; Chalabaev, Sarrazin et al., 2013; Gentile et al., 2018; Smith & Martiny, 2018). Such effects have been noted across technical aspects of many sports—basketball (Krendl et al., 2012), golf-putting (Grimm et al., 2016), soccer (Grabow & Kühl, 2019; Hermann & Vollmeyer, 2016; Heidrich & Chiviacowsky, 2015), and tennis (Hively & El-Alayli, 2014)—and even strength tasks (Deshayes, Clément-Guillotin al., 2019; Deshayes, Zory et al., 2019). As a result, there are now calls to better understand stereotype threat, not merely to further basic research but also to support its application to sporting practice (Smith & Martiny, 2018).

Across all this research, the typical conclusion is that the effects of priming negative stereotypes are almost universally negative. Part of what makes the connection appear so strong is how straightforward it is to produce impairments; a simple word from one person (Hively & El-Alayli, 2014) can trigger the intrapsychic domino chain that leads performance to suffer. However, despite the evidence from this early stereotype threat research, more recent work questions whether the relationship from stereotypes to individual performance is inevitable (e.g., Deshayes, Clément-

Guillotin al., 2019; Deshayes, Zory et al., 2019), or whether targets of social stereotypes can avoid the effects of stereotype activation.

Researchers have developed and tested a number of promising interventions to combat stereotype threat effects. For example, performance seems to be reliably buffered when targets affirm a personal value (Salles et al., 2016), cultivate in-group role models (Van Camp et al., 2019), or practice "routinizing" the basic operations necessary for performance (Beilock et al., 2007). Unfortunately, what nearly all interventions have in common is that they are difficult for non-experts to administer quickly "field-side" (for an exception, see Martiny et al., 2015). The self-affirmation protocol, for example, requires several minutes of writing, which would not be easy to administer on the field or even in a locker room. Math practice works best if it is spaced out over several intensive sessions; the protocol developed by Beilock et al. (2007, Experiment 4) requires three different practice sets of 212 math problems. Cultivating role models can take months or years of in-person interaction—even in its simplest incarnation, it is not something an athlete can plausibly pause to do before a high-stakes performance opportunity. These interventions may be effective, but the time required to implement them makes them impractical.

More importantly, to date, none of these interventions has taken the approach of encouraging targets to explicitly question the validity and legitimacy of negative stereotypes. The fact that such an intervention approach has been overlooked is surprising, given that a number of influential theoretical perspectives identify the benefits of mobilizing an alternative mentality: tapping into challenge rather than threat (Blascovich & Tomaka, 1996); promotion rather than prevention (Higgins, 1998); or resistance to, rather than acceptance of, inequalities and injustices (see Salvatore et al., in press, for a review in the domain of sport). Research informed by the social identity approach to intergroup relations (Tajfel & Turner, 1979) is of particular relevance here because it shows that the motivation to resist or de-legitimate an existing status hierarchy arises spontaneously among members of

disadvantaged groups (Haslam & Reicher, 2012) and that this motivation can be harnessed to improve individual performance (Green et al., 2018; Rees et al., 2013). In summary, other interventions may be overlooking the active and agentic role that targets themselves can play in neutralizing stereotype threat and buffering their own performance.

Cultivating a questioning mindset—one characterized by explicit skepticism of the legitimacy and validity of stereotypes, and critical thinking about the basis of self-relevant stereotypes—seems to have potential, then, as an intervention for stereotype threat. In principle, the benefit could emerge quickly via a reminder that short-circuits the processes that underlie performance impairment. Surprisingly, this idea does not seem to have been directly tested. Only two papers appear to exist that test similar ideas empirically. The first (Alter et al., 2010) shows that reframing stereotype-relevant tasks as challenges rather than threats can combat stereotype threat effects. However, in this research, participants were not asked to directly question the basis for the stereotype and explicitly question its validity and legitimacy. The challenge manipulation was focused on the difficulty or threat of the task, rather than being focused on calling the stereotype itself into question. The second (Smith & Postmes, 2011) is more closely aligned to the present study's manipulation; it shows that challenging a stereotype through an extended group discussion can reduce the extent to which individuals' later behavior conforms to that stereotype. More specifically, the average math performance of men diverges from that of women after a lengthy discussion affirming a well-known stereotype linking gender with math, whereas the groups' performance converges after a lengthy discussion challenging it. It is not clear, however, whether these findings were contingent upon the group processes that emerged in the extended discussion (the length of which is central to the authors' argument), or can be observed much more quickly when individual participants question a stereotype after only minimal cueing from another person.

Here—in the first study of its type—we developed and tested a novel, direct, swift, and real-to-

life stereotype-questioning intervention among several samples of athletes threatened by different stereotypes about athletic competence: female cricketers and female soccer players thinking about threatening gender stereotypes (Experiment 1), male cricketers thinking about threatening stereotypes specific to their team (Experiment 1), and male soccer players thinking about threatening national stereotypes (Experiments 2 and 3). In addition, in Experiment 3, we sought to examine potential processes understood in the stereotype threat literature to mediate poor performance (DeCaro et al., 2011).

Pilot Testing

As all participants in our experiments were based in the United Kingdom (UK), we undertook pilot testing to confirm widespread awareness of several stereotypes within the UK. The approach we took in this regard was similar to that used by others (e.g., Gibson et al., 2014; Martiny et al., 2015) to assess awareness of stereotypes.

First, students studying sport and health sciences (n = 23) were asked to agree or disagree that "Regardless of what I myself believe, most people would privately agree" that women are inferior to men at football (the term for soccer in the UK), at cricket, and at sports in general. The findings confirmed that, in the UK, the stereotype that women are less competent than men at sports is broadly familiar: pilot participants unanimously agreed that most people would privately agree that "women are not as good as men at" both sports, and 19 of the 23 agreed about the statement involving sports in general. Given the UK national dialogue around sexism in soccer (see Magowan, 2011, for an example of part of this dialogue), this was a particularly promising context in which to seek further evidence of our ideas about how salient stereotypes can be effectively neutralized.

Similarly, the same undergraduates were asked to agree or disagree with three statements assessing the stereotype that English footballers are more likely than players of other nationalities to "choke under pressure" when taking penalties. We used the same stem—"Regardless of what I myself

believe, most people would privately agree that"—and then three items: "Compared to the rest of the men's football world, England are terrible at penalty shootouts" (resulting in unanimous agreement); "Other teams in men's world football know that England can be counted upon to choke in penalty shoot-outs" (16 agreed, 6 disagreed, 1 missing data point); and "It must be something about the international spotlight, but English male players just can't cut it in football penalty shootouts" (13 agreed, 10 disagreed). In each item, then, there was at least majority agreement with the idea that a national stereotype about choking under pressure is familiar to most people in the UK context (Burt, 2012)—though note that there was also a non-trivial level of disagreement for two of the three items. Combining with these pilot data are several other sources of information—popular media (e.g., BBC, 2017; Lyttleton, 2014), online discussion forums (e.g., Dernoncourt, 2014), and scholarly publications (e.g., Jordet, 2009)—that together give us reasonable confidence in the familiarity of these stereotypes within the UK context. We used these pilot testing materials to create the targeted stereotype reminders used for each of the samples in these experiments (see Appendix A).

Experiment 1

We first aimed to test the basic hypothesis that questioning stereotypes effectively buffers different types of athletes from the performance impairments induced by various stereotype threats. In addition, we aimed to compare our intervention to an alternative one with established efficacy, namely cognitive load. Stereotype threat works partly by disrupting automatic processes (such as well-learned mental or sporting tasks) in favor of controlled processes (such as actively monitoring our performance; Chalabaev, Brisswalter et al., 2013; Mazerolle et al., 2012; Schmader et al., 2008). To add cognitive load is to increase the burden on working memory. In the case of mental tasks, this subtracts from the resource that needs to be devoted to the task (e.g., math problems), but in the case of sporting tasks, it diverts the athlete's attention away from their motor actions and thereby short-circuits the tendency to overthink those actions under threat. For our samples of high-level athletes, the experimental tasks are

well-learned and automatic motor procedures, so cognitive load should improve their performance under stereotype threat. Importantly, though, we expected that questioning the stereotypes relevant to their identity would be *even more* beneficial. This expectation of our questioning manipulation improving performance over and above the effect of cognitive load necessitated the order of treatment presentation (see below).

For this experiment, we used a multi-phase procedure in which all participants received all three levels of our treatment—first the threat induction, followed by the cognitive load induction, and last the questioning induction—and each phase was followed by a performance opportunity. Given the limited numbers of high-level athletes available as participants, this within-participants design allowed us to conserve participant resources and optimize power. We predicted that cognitive load would improve performance relative to the initial threat, and we further predicted that questioning the stereotypes would further improve performance. To test these predictions, we used repeated-measures analyses of variance to assess changes in performance across the three rounds. We report follow-up comparisons between each round using the Bonferroni adjustment to manage multiple comparisons.

Method

Two types of athletes, soccer (football) players and cricketers, volunteered to participate. All experimental protocols, both here and in subsequent studies, were reviewed and approved by university ethics committees. The soccer sample, comprising 40 members of three teams of female soccer players, aged 18-31, was relatively elite; for example, it included three international-level players. The cricket samples comprised a total of 40 players (20 men aged 18-26 and 20 women aged 18-22), all drawn from strong cricket teams with several players competitive at the regional level. Because of the high player quality in Experiment 1—the required skills were well learned and substantively rehearsed, and the players were both experienced and technically expert—we reasoned that the risk of practice

effects in this experiment—that is, improvement over the course of the study due to skill development—was very low.

All participants in Experiment 1 were tested individually during the course of a normal practice session, in which they had already warmed up and practiced the various component skills of their sport. They were each pulled aside individually for three performance rounds appropriate to their sport—that is, penalty shots in soccer and bowling in cricket.

A penalty shot consists of kicking at the goal while a goalkeeper is present and actively attempting to defend the goal. Our soccer goal was FIFA-approved (2.44m high x 7.32m wide). Each participant took the kick from the penalty spot at a distance of 11m away from the goal, using a size 5 Mitre Calcio soccer ball. A Panasonic NV-M40 video camera was used to confirm the success of the shots. The soccer players took three rounds of ten penalty shots, with 1 scored for each successful penalty and 0 for penalties that were missed or saved by the goalkeeper (three goalkeepers were used, each participating in all three conditions within the group of players they faced; hence goalkeeper is not confounded with condition). Therefore, for each participant we recorded a score out of 10 representing the number of successful penalties at each phase.

Bowling in cricket is similar to baseball pitching, but normally entails the ball bouncing off the ground before it reaches its target. To bowl, participants aimed at a designated target that was 22 yards away (the standard length of a cricket wicket) that we had designed especially for this study. It consisted of a regular set of "stumps" and different target areas around the stumps. The cricketers completed three rounds of a bowling task. They were assigned a score of between 0 and 10 for each bowl, for a total of between 0 and 100 points per round. Participants were given points for where the ball hit the target: ten points if the participant hit the stumps; five if the ball marginally missed the stumps to the left, right or above; two, one or zero as the ball moved further away from the stumps. A video camera with a tripod recorded where each ball had hit the target.

All of the inductions involved a short paragraph spoken out loud by the experimenter to the individual participant. First, we simply reminded participants of the relevant stereotype. The stereotype we used among the men involved the reputation and status of their team within the league. The stereotype we used among the women was a gender stereotype. We designed the threat induction wordings based on familiarity with manipulations used in the stereotype threat literature (e.g., Cadinu et al., 2005; Stone & McWhinnie, 2008). Next, to create cognitive load, following past researchers (e.g., Wegner et al., 1998), we requested that participants take their penalty shots or bowls while counting backwards continuously from one thousand by threes. Lastly, we encouraged participants to question the validity and legitimacy of the stereotype (see Appendix A for details of the wordings we used for the threat inductions and to encourage questioning in each sample).

Results and Discussion

As predicted, the repeated measures ANOVA (with Greenhouse-Geisser correction) revealed that the soccer players' scores improved notably across the three rounds, $F(1.72, 66.95) = 5.42$, $p = .009$, $\eta^2 = .12$. Specifically, they averaged 5.25 shots ($SE = .35$) out of 10 after being reminded of the stereotype that women are inferior athletes, improved to 5.85 shots ($SE = .29$) under cognitive load; and further improved to 6.13 shots ($SE = .28$) after questioning the stereotype. With the Bonferroni correction, neither of the two incremental increases was independently significant (both ps >.07), but the net increase across both phases was ($p = .032$; $d = .42$). Overall, then, participants made nearly a full shot more after they had questioned the validity of the stereotype than when they first were reminded of it.

Independently, the cricketers' scores also improved notably across the three rounds, $F(1.57, 61.02) = 62.76$, $p < .001$, $\eta^2 = .62$. Specifically, they averaged a score of 34.83 ($SE = 1.99$) out of a possible 100 points after being reminded of the negative stereotype; improved to 45.38 ($SE = 2.09$) under cognitive load ($p <.001$ for the incremental improvement); and further improved to 51.03 ($SE =$

2.05) after questioning the stereotype (p <.001 for the incremental improvement). Thus, the questioning process was effective in improving performance over and above the effects of cognitive load.[1] The significant change across both phases (p < .001) was associated with an effect size of d = 1.44.

In sum, as predicted, questioning the stereotype improved mean performance relative to the initial threat. Among the two samples of cricketers, we observed initial improvement under load followed by even greater improvement after questioning the stereotype, and the effect size for the total improvement was considerable. This provides important initial evidence that stereotype threat performance effects are not inevitable; instead, they can be ameliorated or eliminated by a short, simple intervention that includes encouraging athletes to proactively question the validity and legitimacy of the stereotype.

Experiment 2

We recruited a new sample of high-level athletes in order to conduct a test of replication. We also aimed to extend our investigation beyond the prior experiment in two important ways. First, given that cognitive load on its own did not always result in improvements relative to the initial threat in Experiment 1, we omitted further exploration of cognitive load to focus more clearly on the key comparison (threat vs. questioning). Second and more important, we took the opportunity to refine the study design to more definitively rule out the alternative explanation that our results were due to order effects. Although it is not very plausible that the athletes participating in Experiment 1 experienced substantial practice effects, given their high level of play, and not very plausible that they warmed up as time passed, given that a warm-up had already occurred during the session, it is possible that improvement might have stemmed from the increasing fatigue of the *goalkeeper*. Therefore we sought in Experiment 2 to clearly eliminate confounding variables with a between-participants design.

Method

Experiment 2 allowed us to compare threat versus questioning in the context of a between-participants design. Male soccer players (n = 98, aged 18 to 23) volunteered to participate. In a balanced design, each participant (n = 49 in each group) took ten penalty shots after being reminded of the negative national stereotype or after being reminded of it and then encouraged to question and resist it. With only a single opportunity to take penalty shots, improvement due to practice was not possible.

Participants in Experiment 2 were members of university intercollegiate competitive (not intramural) teams. While not internationally elite, all participants were experienced soccer players who trained twice weekly and played twice weekly in competitive British Universities and regional leagues. Given the results of the pilot study as reported above as well as media coverage (Burt, 2012; see also Jordet, 2009, for an example of scholarly attention), we can be confident that this sample of English players was aware of the stereotype that English soccer players choke in international competition. It is important to note that, not being members of the English national team themselves, these participants are a step further removed than those in Experiment 1 from the direct self-relevance of the stereotype.

During a normal practice session, having already warmed up and practiced various typical soccer skills, participants took ten penalty shots as described above for Experiment 1. Prior to this, participants received, by random assignment, either a simple reminder of the stereotype (threat condition) or an invitation to question the stereotype. We again assessed performance by assigning each penalty kick a score of either 0 or 1, so that each participant received a score out of 10 (four goalkeepers were used, each participating in the two conditions within the group of players they faced; hence goalkeeper is not confounded with condition).

Results and Discussion

Consistent with the previous findings, participants in the questioning condition (M = 7.82, SE = .20) scored more penalties than those in the threat condition (M = 6.90, SE = .23), $t(96) = 3.05$, p =

.003, Cohen's d = .62. These findings robustly replicated the performance differences first shown in the previous samples, although we should note that we only tested the logic of our between-participants predictions in one sport (rather than using both soccer and cricket samples, as in Experiment 1). Furthermore, they more definitively established that order effects, if present in Experiment 1, could not fully explain those differences. However, given the lack of a 'true' control condition, we could not speak to what produced performance differences. We sought to address this question in Experiment 3.

Experiment 3

Though results across the first two experiments supported the idea that questioning improves performance, neither included a true control condition. We therefore tested our logic among a further sample of athletes—this time, male soccer players who each took five penalty shots after being randomly assigned to either a true control condition or the threat condition or the threat-then-questioning condition. The true control was intended to illuminate whether the effect was more attributable to impairments (under threat) or to improvement (with questioning).

In addition, we compared three performance-undermining mental processes noted in the stereotype threat literature (Beilock & McConnell, 2004; Schmader et al., 2008; Schmader, 2010) that might be helpfully disrupted by the experience of questioning the legitimacy of a stereotype. We asked participants to report on their conscious experiences of monitoring and thinking about their performance (a cognitive mechanism); of automatically executing the movement without active thought (a motor mechanism); and of worrying about "messing up" (an affective mechanism; see also Hermann & Vollmeyer, 2015). We chose these because various processes can mediate poor performance either individually or in combination (DeCaro et al., 2011), and we reasoned that it would be helpful to distinguish between them.

Method

Male soccer players (n = 132, aged 18 to 32) volunteered to participate in Experiment 3. The

procedure was nearly identical to the one we used with Experiment 2, with the following modifications. First, to reduce the length of the session and make it possible for more players to participate, participants took five penalty shots instead of ten. Second, to optimize interpretability, we added a true control condition: by random assignment without replacement, one-third of participants took their penalty shots without hearing anything at all about the national stereotype. In this experiment, two goalkeepers were used, each participating in all three conditions within the group of players they faced; hence goalkeeper is not confounded with condition.

Last, to distinguish between various mechanisms, we assessed several potential mediators with one self-report item each, responded to on a scale ranging from 1 (*not at all*) to 5 (*extremely so*). We assessed a cognitive/attentional mechanism (conscious experiences of monitoring) with the wording *While taking your penalty kicks, to what extent were you deliberately monitoring and thinking about your technique and the process of penalty-taking*? We assessed a motor mechanism (automatically executing the movement without conscious thought) with the wording *While taking your penalty kicks, to what extent were you doing things automatically and without thinking*? We assessed an affective mechanism (worry about making errors) with the wording *While taking your penalty kicks, to what extent were you worried about messing up*? We analyzed mediation using the PROCESS module for SPSS (Hayes & Preacher, 2014).

Results & Discussion

Replicating findings from the previous samples, we observed significant differences in penalty performance across the three groups, $F(2,129) = 11.02$, $p < .001$, $\eta^2 = .15$. The stereotype-threatened group ($M = 2.75$, $SE = .19$) performed worse than the questioning group ($M = 3.50$, $SE = .14$), Tukey's HSD $p = .003$, and also worse than the control group ($M = 3.75$, $SE = .13$), Tukey's HSD $p < .001$. There was no difference between the questioning group and the control group, Tukey's HSD $p = .499$

(see Figure 4). Questioning appears to help protect athletes from the performance impairments that would otherwise be triggered by activation of negative stereotypes.

Descriptive statistics for the mechanisms variables are reported in Table 1. Analyses of variance suggested that there were differences across conditions for these mechanisms variables. For conscious experiences of monitoring, ANOVA ($F_{2,129} = 28.57$, $p < .001$, $\eta^2 = .31$) with Tukey HSD follow-ups demonstrated differences between control and threat ($p < .001$), between control and questioning ($p = .01$) and between threat and questioning ($p < .001$). For automatically executing the movement, ANOVA ($F_{2,129} = 22.142$, $p < .001$, $\eta^2 = .26$) with Tukey HSD follow-ups demonstrated differences between control and threat ($p < .001$), and between threat and questioning ($p < .001$), but not between control and questioning ($p = .07$). For worrying about messing up, ANOVA ($F_{2,129} = 33.20$, $p < .001$, $\eta^2 = .34$) with Tukey HSD follow-ups demonstrated differences between control and threat ($p < .001$), between control and questioning ($p = .002$) and between threat and questioning ($p < .001$). Initial correlational analyses supported a relationship between all three of the potential mechanisms and performance. Conscious experiences of monitoring ($r = -.35$, $p < .001$), automatically executing the movement ($r = .30$, $p < .001$), and worrying about messing up ($r = -.41$, $p < .001$) were each associated with penalty performance. In a regression with all three predictors entered simultaneously, however, only worrying about messing up emerged as a significant predictor on its own ($b = -.25$, $p = .007$).

Next we compared the indirect effects of the three potential mechanisms. First we re-coded the condition variable using indicator coding, with the threat condition providing the reference group. Subsequently, we examined whether all three potential mechanisms, entered simultaneously into one analysis, acted as mediators of the performance effect. Of the three, consistent with the statistic reported in the prior paragraph, only worrying was significantly associated with performance ($b = -.21$, $p = .03$). Furthermore, it was only for the worrying variable that the confidence intervals around the

indirect effects of the questioning group (versus threat: .22, CI [.03, .49]) and the control group (versus threat: .38, CI [.05, .80]) did not include zero, suggesting that it was this mechanism that mediated the performance effect, rather than the other two potential mechanisms. We concluded that the more cognitive and motor mechanisms did not produce the stereotype threat disruption effect.

With a true control condition as a comparison for the interventions, these findings provided further evidence of our claim that directly questioning stereotypes buffers the performance impairment that would otherwise occur when the stereotype was front of mind. It also provides initial evidence of *how* this simple intervention does so (although we should note the psychometric limitations inherent in the use of our single-item process measures). Questioning does not appear to stop athletes from overthinking a familiar and thus typically unmonitored task, and it does not appear to disrupt their execution of that task; rather, it appears to stop them from worrying about "messing up."

General Discussion

Stereotype threat effects have been so extensively documented in sport psychology that the literature has matured into a phase of meta-analyses (e.g., Gentile et al., 2018) and commentary (e.g., Smith & Martiny, 2018). As we noted at the outset, most researchers now accept the conclusion that, unless an active intervention occurs, such effects reliably lead to impaired performance. Such interventions, however, typically require a good deal of time—or at a minimum, a writing implement or a screen—to implement, and they target micro-level intrapsychic processes that have no larger implications for the stereotype itself. The aim of the present research was to contribute to this effort by demonstrating that these impairments can be productively disrupted in a much simpler way. We found that the simple action of inspiring participants to question the validity and legitimacy of a threatening stereotype to which they were exposed was enough to protect them from the performance impairments that one would normally expect to follow. Interestingly, this logic may generalize beyond the domain

of sport into more socio-political arenas. Indeed, questioning can be a tool for the reorganization of entrenched social / socio-political hierarchies—both in sport and beyond.

The effect we observed was robust—it held across tests in multiple samples involving two different sports and a variety of stereotypes—and it is important on both theoretical and practical grounds. Theoretically, we were able to show that resistance to stereotypes need not require extended group conversation (Smith & Postmes, 2011) to be effective. In addition, we were able to provide (in our final experiment) some preliminary evidence about how and why these effects might have occurred. Our observations among the final sample suggest that, rather than our intervention protecting performance via processes of explicit monitoring or motor automaticity, participants were helped via the process (cf. Martiny et al., 2015) of becoming less concerned about making errors.

Practically speaking, too, the finding that effective protection from stereotypes need not require extended multi-person input has important implications when working in the field. Indeed, all those working with athletes will welcome the news that a brief intervention has such a powerful effect (Smith & Martiny, 2018), because most interventions and manipulations typically require a significant amount of time. Indeed, the intervention we tested here is an extremely practical one: in contrast to other intervention strategies (see, e.g., Smith & Postmes, 2011), it can be easily administered by coaches, leaders, managers, and educators "in the moment" and without cost. In its ease of use, it is akin to an intervention in which coaches can remind negatively stereotyped team members (e.g., women athletes) that they are *also* positively stereotyped (e.g., high-ranked within their league; Martiny et al., 2015). Since directly questioning stereotypes is both simple and effective, it deserves to be widely available to practitioners as an intervention option.[2] The above said, one would need to use caution in directly applying the results from these experiments, in cases where athletes had not already been negatively affected by stereotype threat, because the performance of our questioning group (in Experiment 3) was slightly lower (albeit non-significantly) than the control group. The main implication would be that

if/when coaches hear or see athletes experiencing a negative stereotype, they should indeed then use convincing, counter-arguments (i.e., by questioning the stereotype), to protect them from its potential negative effects.

As we noted above, our results held across samples from two sports (soccer and cricket) and with sub-samples of both male and female athletes from each sport; in tests employing both within- and between-group designs; and regardless of the specific content or scale of the stereotype. The key seems to be simply that the association is firmly known by those to whom it pertains; the intervention protected against impairment stemming from content as wide-ranging and varying in scale as gender stereotypes, a national stereotype, and a stereotype about a specific team's reputation and status within its own league of competition.

We also employed reasonably high-level samples of participants (including some international performers), our effect sizes ranged from moderate to large, and, mindful of ecological validity, we conducted all experiments as part of regular training sessions. However, it is also true that in some examples, we used well-skilled student athletes, and our effect sizes were higher when performance was based on our measure of accuracy (in cricket) rather than goals scored (in soccer). These are important points to consider, because the stereotype threat literature generally implies that effects should be the most pronounced among athletes who are most invested—highly-skilled in and identified with their sport as well as with the stereotyped group (Schmader et al., 2008; Smith & Martiny, 2018). Further, it is worth noting that, although the literature in sport has recently flagged the importance of this topic, the wider social psychology literature has begun to question the previously observed large effects. As Shewach and colleagues noted in their recent meta-analysis, stereotype threat effects may only be small, at best—the initially reported effects of much greater magnitude may have been inflated by publication bias. This is not to suggest that this is the case within sport, or to negate the fact that

such small effects "compounded over many individuals and across time, can yield substantial consequences" (Shewach et al., 2019, p. 1), but it does necessitate caution.

Alongside the differences in performance assessment, we should note that some differences between experimental procedures may also have influenced our results beyond those noted earlier (such as the reduction in penalties taken in Experiment 3). The type of grass—real grass versus artificial (rubber crumb) grass—varied across the experiments, although this perhaps speaks more to the generalizability of the results across various conditions. Further, the change in experimenters across the samples might also have influenced the results, because identification with experimenters can influence participants to align their behavior with the goals of research (Haslam et al., 2019). Given the consistent pattern of results across the present series of experiments, however, it is unclear how such identification could explain differences of the form we observed. The role that identification with the experimenter plays in producing performance and process effects in stereotype threat research would nonetheless be worth exploring in future (see also Green et al., 2018). Another concern might be whether the stereotypes we used in the present set of experiments were salient enough to actually be experienced as "threatening" to the athletes. Although we noted our confidence, for example, that the present samples of English soccer players were aware of the stereotype that English soccer players choke in international competition, a measure of salience, agreement, interpretation, or appraisal could have provided further argument for the relevance (and mechanism) of the stereotypes we used. We would add that anyone seeking to further develop this work looking at gender should attend more carefully to the specific language of the manipulation to equalize the strength of the manipulation across stereotype domains and ensure that the wording does not unintentionally prime the stereotype further. And finally, because there was no baseline performance assessment in Experiments 2 and 3, we relied solely on random assignment as a tool for attempting to equalize pre-existing performance variation; future work would do well to attend closely to ensuring this equalization.

In summary, this research reminds us how powerful stereotype threat can be, and presents a timely rejoinder to the message that its negative effects are unavoidable. As the field is beginning to recognize, stereotype threat effects may be powerful, but they are not inevitable (see also Stafford, 2018). Targets of stereotypes can play an active and agentic role in questioning, resisting, and thereby overcoming them.

References

Alter, A. L., Aronson, J., Darley, J. M., Rodriguez, C., & Ruble, D. N. (2010). Rising to the threat: Reducing stereotype threat by reframing the threat as a challenge. *Journal of Experimental Social Psychology*, *46*(1), 166-171. https://doi.org/10.1016/j.jesp.2009.09.014

BBC News. (2017) Reality check: Are England terrible at penalty shootouts? Retrieved from https://www.bbc.com/news/uk-england-40432001

Beilock, S. L., & McConnell, A. R. (2004). Stereotype threat and sport: Can athletic performance be threatened? *Journal of Sport & Exercise Psychology, 26*(4), 597-609. https://doi.org/10.1123/jsep.26.4.597

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General, 136,* 256-276. https://doi.org/10.1037/0096-3445.136.2.256

Blascovich, J., & Tomaka, J. (1996). The biopsychosocial model of arousal regulation. In *Advances in experimental social psychology* (Vol. 28, pp. 1-51). Academic Press, San Diego, CA.

Burt, J. (2012, June 25). Euro 2012: Spot-kick obsession weighing England down, laments Roy Hodgson, after defeat to Italy. *The Daily Telegraph.* http://www.telegraph.co.uk/sport/football/teams/england/9353340/Euro-2012-spot-kick-obsession-weighing-England-down-laments-Roy-Hodgson-after-defeat-to-Italy.html

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science*, *16*(7), 572-578. https://doi.org/10.1111/j.0956-7976.2005.01577.x

Chalabaev, A., Brisswalter, J., Radel, R., Coombes, S. A., Easthope, C., & Clément-Guillotin, C. (2013). Can stereotype threat affect motor performance in the absence of explicit monitoring processes? Evidence using a strength task. *Journal of Sport & Exercise Psychology*, *35*(2), 211-215. https://doi.org/10.1123/jsep.35.2.211

Chalabaev, A., Sarrazin, P., Fontayne, P., Boiché, J., & Clément-Guillotin, C. (2013). The influence of sex stereotypes and gender roles on participation and performance in sport and exercise: Review and future directions. *Psychology of Sport and Exercise, 14*(2), 136-144. https://doi.org/10.1016/j.psychsport.2012.10.005

DeCaro, M. S., Thomas, R. D., Albert, N. B., & Beilock, S. L. (2011). Choking under pressure: Multiple routes to skill failure. *Journal of Experimental Psychology: General*, *140*(3), 390-406. https://doi.org/10.1037/a0023466

Dernoncourt, F. (2014, July 17). Football penalty shootouts: Germany wins 71% of the time, England 17% only. Any reason? [Online forum post]. Stack Exchange Network: Sports. https://sports.stackexchange.com/questions/5349/football-penalty-shootouts-germany-wins-71-of-the-time-england-17-only-any

Deshayes, M., Clément-Guillotin, C., & Zory, R. (2019). "Men are better than women!" The positive effect of a negative stereotype toward women on a self-paced cycling exercise. *Journal of Sport & Exercise Psychology, 41*(4), 242-250. https://doi.org/10.1123/jsep.2018-0213

Deshayes, M., Zory, R., Seitchik, A. E., Chalabaev, A., & Clément-Guillotin, C. (2019). Can the stereotype threat and lift phenomenon be applicable to a muscular endurance task? *Research Quarterly for Exercise and Sport*. https://doi.org/10.1080/02701367.2019.1668518

Gentile, A., Boca, S., & Giammusso, I. (2018). 'You play like a woman!' Effects of gender stereotype threat on women's performance in physical and sport activities: A meta-analysis. *Psychology of Sport and Exercise, 39,* 95–103. https://doi.org/10.1016/j.psychsport.2018.07.013

Gibson, C. E., Losee, J., & Vitiello, C. (2014). A replication attempt of stereotype susceptibility (Shih, Pittinsky, & Ambady, 1999). *Social Psychology, 45*(3), 194-198. http://dx.doi.org/10.1027/1864-9335/a000184

Grabow, H., & Kühl, M. (2019). You don't bend it like Beckham if you're female and reminded of it: Stereotype threat among female football players. *Frontiers in Psychology, 10:*1963. https://doi.org/10.3389/fpsyg.2019.01963

Green, J., Rees, T., Peters, K., Sarkar, M., & Haslam, S. A. (2018). Resolving not to quit: Evidence that salient group memberships increase resilience in a sensorimotor task. *Frontiers in Psychology, 9:2579*. https://doi.org/10.3389/fpsyg.2018.02579

Grimm, L. R., Lewis, B., Maddox, W. T., & Markman, A. B. (2016). Stereotype fit effects for golf putting nonexperts. *Sport, Exercise, and Performance Psychology, 5*(1), 39–51. https://doi.org/10.1037/spy0000047

Haslam, S. A. & Reicher, S. D. (2012). When prisoners take over the prison: A social psychology of resistance. *Personality and Social Psychology Review, 16*(2), 154-179. https://doi.org/10.1177/1088868311419864

Haslam, S. A., Reicher, S. D., & Van Bavel, J. J. (2019). Rethinking the nature of cruelty: The role of identity leadership in the Stanford Prison Experiment. *American Psychologist, 74*(7), 809–822. https://doi.org/10.1037/amp0000443

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, *67*(3), 451-470. https://doi.org/10.1111/bmsp.12028

Heidrich, C., & Chiviacowsky, S. (2015). Stereotype threat affects the learning of sport motor skills. *Psychology of Sport and Exercise, 18,* 42-46. http://dx.doi.org/10.1016/j.psychsport.2014.12.002

Hermann, J. M., & Vollmeyer, R. (2016). "Girls should cook, rather than kick!" Female soccer players under stereotype threat. *Psychology of Sport and Exercise, 26,* 94-101. http://dx.doi.org/10.1016/j.psychsport.2016.06.010

Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. In *Advances in experimental social psychology* (Volume 30, pp. 1-46). Academic Press, San Diego, CA.

Hively, K., & El-Alayli, A. (2014). "You throw like a girl": The effect of stereotype threat on women's athletic performance and gender stereotypes. *Psychology of Sport & Exercise, 15*(1), 48–55. https://doi.org/10.1016/j.psychsport.2013.09.001

Jordet, G. (2009). Why do English players fail in soccer penalty shootouts? A study of team status, self-regulation, and choking under pressure. *Journal of Sports Sciences*, *27*(2), 97-106. https://doi.org/10.1080/02640410802509144

Krendl, A., Gainsburg, I., & Ambady, N. (2012). The effects of stereotypes and observer pressure on athletic performance. *Journal of Sport & Exercise Psychology, 34*(1), 3-15. https://doi.org/10.1123/jsep.34.1.3

Lyttleton, B. (2014, June 6). England: Why are they so bad at penalties? *Financial Times Magazine*. https://www.ft.com/content/f32c2ac6-ec4a-11e3-ab1b-00144feabdc0

Magowan, A. (2011, January 24). Is sexism still rife in football? *The BBC*. http://news.bbc.co.uk/sport2/hi/football/9371621.stm

Martiny, S. E., Gleibs, I. H., Parks-Stamm, E. J., Martiny-Huenger, T., Froehlich, L., Harter, A., & Roth, J. (2015). Dealing with negative stereotypes in sports: The role of cognitive anxiety when multiple identities are activated in sensorimotor tasks. *Journal of Sport & Exercise Psychology*, *37*(4), 379-392. https://doi.org/10.1123/jsep.2014-0284

Mazerolle, M., Régner, I., Morisset, P., Rigalleau, F., & Huguet, P. (2012). Stereotype threat strengthens automatic recall and undermines controlled processes in older adults. *Psychological Science*, *23*(7), 723-727. https://doi.org/10.1177/0956797612437607

Rees, T. J., Salvatore, J., Coffee, P., Haslam, S. A., Sargent, A., & Dobson, T. (2013). Reversing downward performance spirals. *Journal of Experimental Social Psychology, 49*, 400-403. https://doi.org/10.1016/j.jesp.2012.12.013

Salles, A., Mueller, C. M., & Cohen, G. L. (2016). A values affirmation intervention to improve female residents' surgical performance. *Journal of Graduate Medical Education*, *8*(3), 378-383. https://doi.org/10.4300/JGME-D-15-00214.1

Salvatore, J., Sankaran, S., & Cole, D. (2020). Cognition and sporting performance. In S. A. Haslam, K. Fransen, & F. Boen (Eds.), *The new psychology of sport and exercise: The social identity approach* (pp. 111-128). London: Sage. https://uk.sagepub.com/en-gb/eur/the-new-psychology-of-sport-and-exercise/book265588

Schmader, T. (2010). Stereotype threat deconstructed. *Current Directions in Psychological Science*, *19*(1), 14-18. https://doi.org/10.1177/0963721409359292

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

Shewach, O. R., Sackett, P. R., & Quint, S. (2019). Stereotype threat effects in settings with features likely versus unlikely in operational test settings: A meta-analysis. *Journal of Applied Psychology, 104*(12), 1514–1534. https://doi.org/10.1037/apl0000420

Smith, D. M., & Martiny, S. E. (2018). Stereotype threat in sport: Recommendations for applied practice and research. *The Sport Psychologist, 32*(4), 311-320. https://doi.org/10.1123/tsp.2017-0134

Smith, L. G. E., & Postmes, T. (2011). Shaping stereotypical behaviour through the discussion of social stereotypes. *British Journal of Social Psychology*, *50*(1), 74-98. https://doi.org/10.1348/014466610X500340

569 Stafford, T. (2018). Female chess players outperform expectations when playing men. *Psychological*
570 *Science*, *29*(3), 429-436. https://doi.org/10.1177/0956797617736887
571 Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of
572 stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in experimental social*
573 *psychology* (Vol. 34, pp. 379-440). San Diego, CA: Academic Press.
574 Stone, J., & McWhinnie, C. (2008). Evidence that blatant versus subtle stereotype threat cues impact
575 performance through dual processes. *Journal of Experimental Social Psychology*, *44*(2), 445-
576 452. https://doi.org/10.1016/j.jesp.2007.02.006
577 Stone, J., Perry, W., & Darley, J. M. (1997). "White men can't jump": Evidence for the perceptual
578 confirmation of racial stereotypes following a basketball game. *Basic and Applied Social*
579 *Psychology, 19*(3), 291-306. https://doi.org/10.1207/s15324834basp1903_2
580 Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S.
581 Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-47). Monterey, CA:
582 Brooks-Cole.
583 Van Camp, A. R., Gilbert, P. N., & O'Brien, L. T. (2019). Testing the effects of a role model
584 intervention on women's STEM outcomes. *Social Psychology of Education*, *22*(3), 649-671.
585 https://doi.org/10.1007/s11218-019-09498-2
586 Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology,*
587 *39*(5), 456–467. https://doi.org/10.1016/S0022-1031(03)00019-2
588 Wegner, D. M., Ansfield, M., & Pilloff, D. (1998). The putt and the pendulum: Ironic effects of the
589 mental control of action. *Psychological Science*, *9*(3), 196-199. https://doi.org/10.1111/1467-
590 9280.00037
591

Footnotes

[1] To rule out any potential differential effects between male and female cricketers (and given their differential primed stereotypes), we also performed a mixed model (2: male/female by 3: phases) analysis of variance. In this analysis, there was a significant main effect for phases ($F_{1.51, 57.49} = 61.29$, $p < .001$), no main effect for gender ($F_{1, 38} = 1.07$, $p = .31$), and no interaction between gender and phases ($F_{1.51, 57.49} = .09$, $p = .86$). The latter non-significant interaction implies that there was no differential effect across phases between male and female cricketers.

[2] Although it is not central to our focus in this article, it is worth noting that cognitive load also appears to be an effective intervention, and thus further research here is warranted.

Table 1

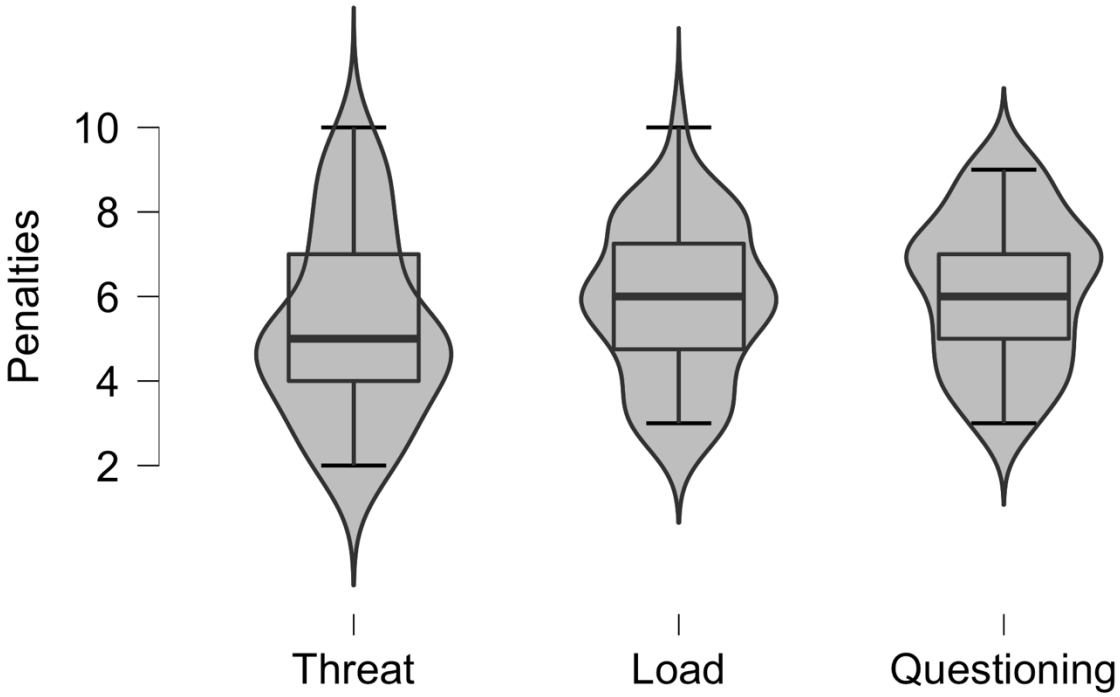
Means (SD) for mechanisms across the three conditions

Condition	Monitoring	Automaticity	Worry
Control	2.09 (.94)	3.75 (.92)	1.93 (.95)
Threat	3.57 (.93)	2.36 (1.08)	3.73 (1.13)
Questioning	2.66 (.94)	3.27 (1.02)	2.71 (1.05)

Note. Monitoring = conscious experiences of monitoring; Automaticity = automatically executing the movement; Worry = worrying about messing up

Figure 1

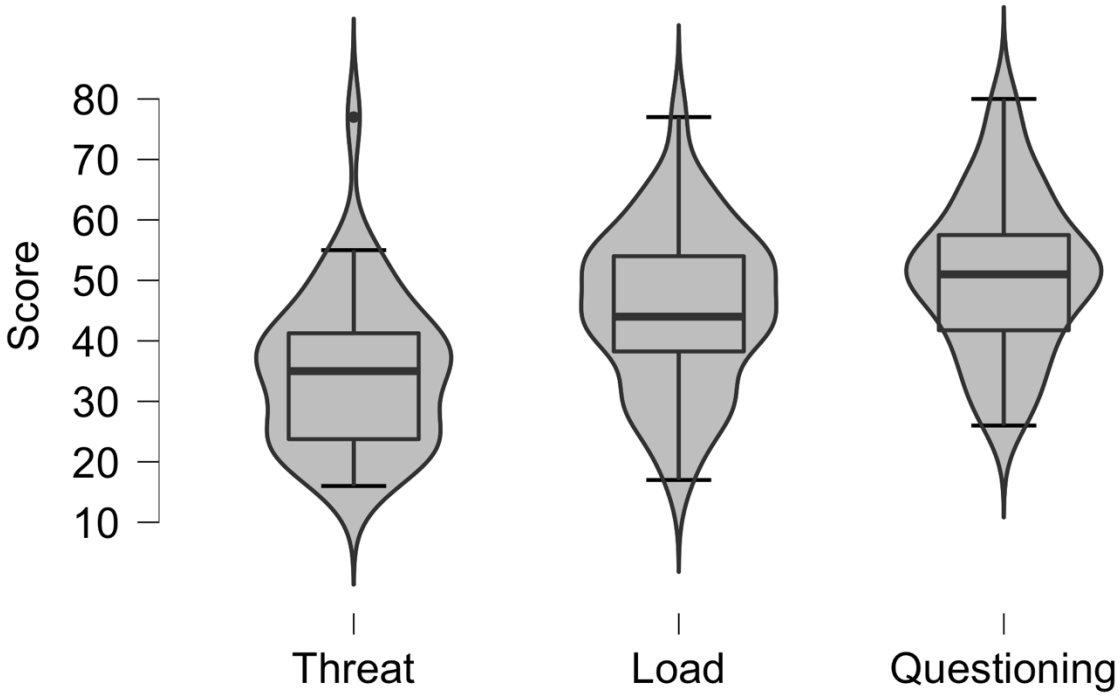
Changes in Soccer Players' Mean Penalty Kick Performance across the three Phases in Experiment 1.



Note: *Soccer players' performance improved significantly from the threat to the questioning phase.*

Figure 2

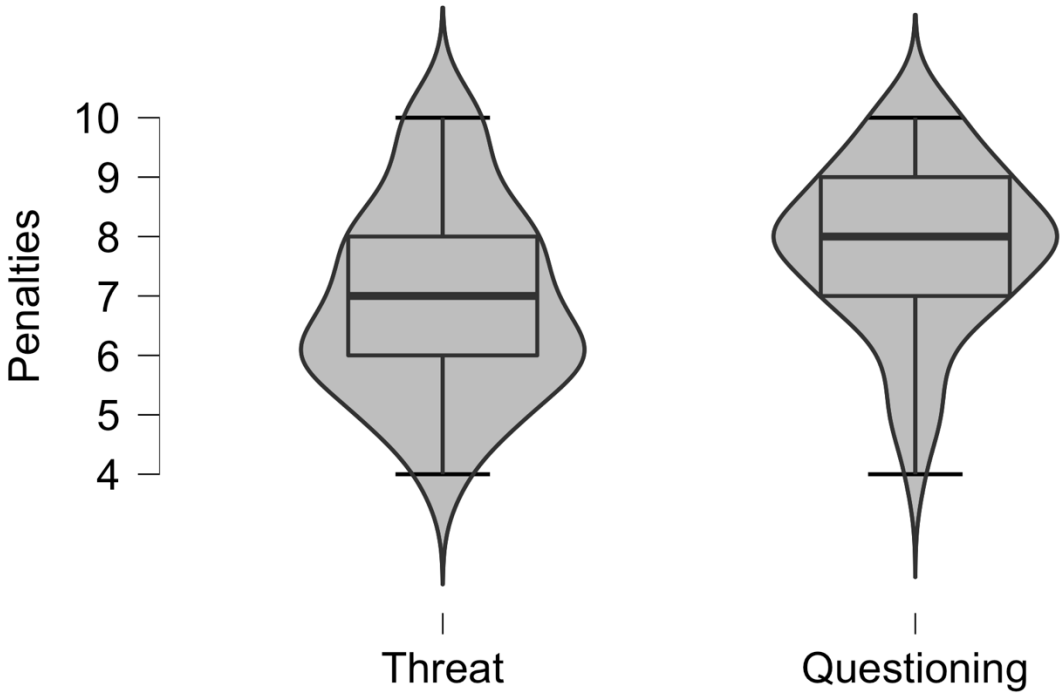
Changes in Cricketers' Mean Bowling Performance across the three Phases in Experiment 1.



Note: *Cricketers' performance improved across the three phases: Each performance increment was independently significant.*

Figure 3

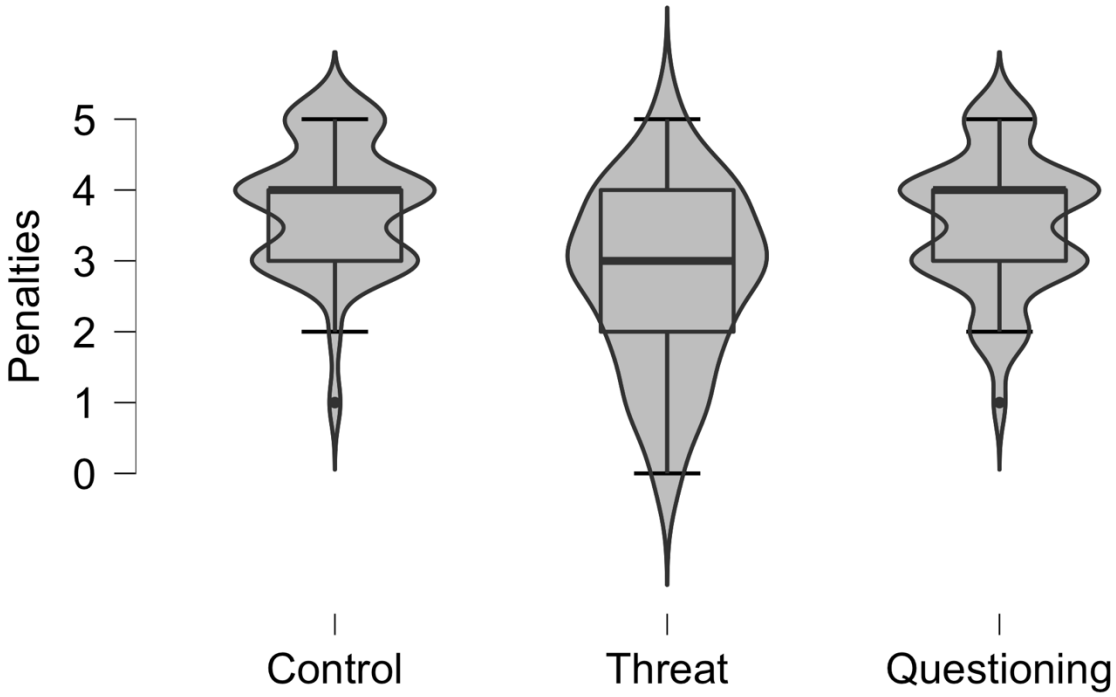
Differences in Soccer Players' Mean Penalty Kick Performance between Conditions in Experiment 2.



Note: Soccer players scored significantly more penalties in the questioning condition than in the threat condition.

Figure 4

Differences in Soccer Players' Mean Penalty Kick Performance between Conditions in Experiment 3.



Note: Soccer players scored significantly fewer penalties in the threat condition than in the control and questioning conditions.

631 **Appendix A**

632 *Threat and Questioning Scripts across Experiments*

	Threat	**Questioning**
Female Soccer Players	"It is the perception of the general public, that women are not as good as men at sport. Women's football in particular is seen as slow and less technical, which may account for the little media attention it receives in comparison with men's. It is felt by many that women are less talented at football, especially in technical skills, such as penalty-taking."	"As you know, it's completely crazy to think that men are better than women at playing football or even taking penalties. Women can be amazing athletes and are at least as good as men. The way they fail to get respect is shocking."
Female Cricketers	"As you probably know, many people think that women are not as good as men at sport, especially when it comes to cricket. There's very little coverage of women's cricket on TV and nobody wants to pay to go and watch it as they don't expect a high quality game. A lot of people think that men are much more talented at cricket than women, especially in terms of their technical ability in bowling."	"As you know, it's ridiculous to think that men are much better at cricket than women, especially when it comes to bowling. Women may not be able to bowl quite as fast as men, but that doesn't mean that they have less technical ability or are less accurate. There are loads of great bowlers in women's cricket—just look at the England Ladies team! It's shocking that women don't get the respect they deserve in cricket."
Male Cricketers	"As you know, many people think that we don't deserve to be in this league. For example, it's well known that, although we got promoted into this league before, we went straight back down the next season, losing every game. So, we're just not taken as seriously as other teams, and no-one expects a high quality game from us. Speaking to some of the lads from [rival team], they certainly don't believe that we're capable of winning a game in this league and they see us as an easy 20 points this season. Teams believe that we are nowhere near as talented at cricket as they are and our bowling attack is the specific weakness."	"As you know, it's completely crazy to think that we don't deserve to be in this league. There's no reason to think we should go straight back down again this season. Some might not think we should be taken seriously, but if that's what they think, they're going to be in for a shock, as we will be bringing a high quality game with us. We have been training hard as a team, we're all three years older and more experienced, and also we have [player name] back this year and other teams are afraid of him! We also have our own pitch again this season, as last time it was ruined due to the floods, and we all know that home advantage is a great benefit; nobody will want to play at [name of town]! As last season showed, this league is where we belong and the way we fail to get the respect we deserve is just shocking. It's going to be different this season."
Male Soccer Players	"Okay, so in this experiment you're being tested on your ability to take penalties. I'm doing this because as you know, many people think that English footballers are terrible at taking penalties. We've lost something like 7 out of the last 8 penalty shootouts in international tournaments and I think other teams know we can be counted upon to choke in these situations. It must be something about the English mentality, but we just don't seem to be able to handle the pressure of penalties"	"Okay, so in this experiment you're being tested on your ability to take penalties. I'm doing this because as you know, some people think that English footballers are naturally bad at taking penalties, but that is completely ridiculous. It all stems from a run of bad luck at penalty shootouts in international tournaments but that's got nothing to do with talent. Our national players score penalties all the time when they are playing for their club teams and we're pretty good at shooting in general play. I think the way we fail to get respect is completely undeserved".

633