Journal of Experimental Social Psychology 92 (2021) 104059

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Math approach training changes implicit identification with math: A close preregistered replication☆

Cédric Batailler[a], Dominique Muller[a,b,*], Cécile Nurra[c], Marine Rougier[d], David Trouilloud[e]

[a] Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, LIP/PC2S, F-38000 Grenoble, France
[b] Institut Universitaire de France, 1 rue Descartes, 75005 Paris, France
[c] Univ. Grenoble Alpes, LaRAC, F-3800 Grenoble, France
[d] Université catholique de Louvain, Belgium
[e] Univ. Grenoble Alpes, SENS, F-38000 Grenoble, France

ARTICLE INFO

Keywords:
Approach and avoidance training
Implicit identification
Preregistered replication
Evidential value
VAAST

ABSTRACT

Kawakami et al. (2008) offer an approach training intervention to strengthen women's implicit identification with math. This intervention is especially interesting regarding data suggesting that the low implicit identification with math among women and the gender gap in science reinforce each other (Nosek et al., 2009). Nevertheless, Kawakami et al.'s data only provide quite modest evidential value in favor of the effectiveness of this intervention (notably because of the two critical p values being very close to 0.05 and of the small sample sizes). In the present manuscript, we offer a preregistered replication of Kawakami et al. with a substantially larger sample size and a novel implementation of approach and avoidance (the VAAST; Rougier et al., 2018). In a Pilot Experiment ($N = 150$), we validate the VAAST-based approach/avoidance training as a way to create identification for novel stimuli ($d_s = 1.17, p < .001$). We then replicate Kawakami et al.'s work, revealing that women who approached math (instead of avoiding it) had a higher identification with math ($d_s = 0.30$, $p = .037$). This preregistered replication increased evidential value in favor of the effectiveness of the approach training by a factor of 2.57, now providing "strong" support for the effectiveness of the training (Jeffreys, 1961). A meta-analysis of the original data and the replication revealed a small-to-medium effect of this intervention ($d_s = 0.40$, $CI_{95\%}$ [0.14; 0.65]). These results are discussed in regard to theories explaining how actions affect evaluative response as well as current interventions in the literature.

1. Introduction

Implicit identification is theorized as the determinant of spontaneous tendencies to associate oneself with concepts, people, or stimuli (Greenwald et al., 2002). Besides mediating the relationship between a stimulus and its indirect evaluation (see De Houwer et al., 2013), for example with the identification Implicit Association Test (identification IAT; Nosek et al., 2002), implicit identification also mediates the relationship between a concept and some of its related behavior. Therefore, to address problematic behaviors, researchers have been trying to directly target implicit identification.

One important area where social psychology researchers have been investigating the role of implicit identification is the gender gap in science (Wang & Degol, 2017). A large body of work has explored the role of identity-related processes in the under-representation of women in STEM-related fields, showing, for example, that a salient female identity could impact women's performance in math (Spencer et al., 1999), even among lower elementary grades (Ambady et al., 2001). In a large-scale study, Nosek et al. (2009) went further by examining an indirect evaluation intended to measure implicit associations between male (female) and science (liberal arts) at the country level. Using data from the 2003 Trends in International Mathematics and Science Study (Gonzales et al., 2004), Nosek et al. were able to predict countries' gender gap in math achievement using scores from the Implicit Association Test (IAT; Greenwald et al., 1998): The stronger the male-science association, the bigger the gender gap. The conclusion they drew out of this correlation was that the bias to associate math more with males than with females and gender gap in math achievement would reinforce each other. That is, low representation of women in science would make it harder for women to identify with math and this dissociation would drive women away from math, ultimately making women underrepresented in this field. In such a context, making

☆ This paper has been recommended for acceptance by Nicholas Rule.

* Corresponding author at: LIP/PC2S, Univ. Grenoble Alpes, CS 40700, 38 058 Grenoble Cedex 9, France.
E-mail address: dominique.muller@univ-grenoble-alpes.fr (D. Muller).

https://doi.org/10.1016/j.jesp.2020.104059
Received 7 August 2019; Received in revised form 3 September 2020; Accepted 13 September 2020

identification between women and math higher would create a positive feedback loop reducing the gender gap in science. This idea makes any intervention reducing the dissociation between women and math extremely valuable. In the present manuscript, we offer a close preregistered replication of such an intervention using an approach and avoidance training (Kawakami et al., 2008).

2. Increasing math implicit identification with an approach training

In an influential paper (more than 90 citations according to a Google Scholar query), Kawakami et al. (2008) presented two studies investigating whether merely approaching math-related pictures could change women's math identification. Recruiting women who reported not liking math, they demonstrated that repeatedly approaching (avoiding) math-related pictures and avoiding (approaching) arts-related pictures had an impact on identification and attitude IATs (Greenwald et al., 1998; Nosek et al., 2002). In these studies, participants approached and avoided stimuli by respectively pulling a joystick (as if they were bringing an object toward them) and pushing it (as if they were pushing away an object). Approaching math (instead of avoiding it) reinforced the participants' self-identification with math and made them evaluate math more positively. We think this paper offers an important leverage to reduce the gender gap in science.

Now, with the goal to build upon this theoretical foundation, we first need to assess evidential value for this effect, that is the strength of evidence that has been collected to support it. One classic way to do so when enough studies have been conducted is to rely on a formal meta-analysis (even though there have been discussions about the efficiency to correct for the publication bias, see Carter et al., 2019—the tendency to publish mostly significant findings—and for degrees of freedom in data analysis when studies have not been preregistered; Simmons et al., 2011). Because, to the best of our knowledge, there is no work other than Kawakami et al.'s (2008) testing the math approach training effect, a formal meta-analysis is not an option. Fortunately, the literature is providing more and more tools to assess evidential value when only few studies are available.

3. Assessing the evidential value in favor of the math approach training

To begin with, and somewhat at odds with an all or nothing interpretation of *p* values (i.e., significant or not with regard to the 0.05 threshold; Giner-Sorolla, 2016), one can often infer stronger evidential value when significant *p* values are farther away from 0.05 (that explains that, for instance, Benjamin et al., 2018 suggested a 0.005 threshold for stronger evidence). Along the same line, Simonsohn et al. (2014) developed an index (the *p*-curve) that relies on the principle that when an effect exists in the population, smaller significant *p* values should be more frequent than significant *p* values close to 0.05 (i.e., a downward slope). Such a tool, however, is not suited to assess evidence from a single paper (Simonsohn et al., 2014). Beyond *p* values, information that contributes to the assessment of evidential value includes effect sizes (i.e., whether most of the effect sizes are of the same direction and how far they are from the null effect), sample sizes (i.e., studies with more observations have less chance to reflect Type 1 errors; Lakens & Evers, 2014; Schönbrodt & Perugini, 2013), and the precision of an effect size (i.e., how narrow a confidence interval is, a narrower one coming with more confidence regarding the existence of an effect—provided obviously that 0 is not included in this confidence interval). It is worth mentioning that methodological aspects can also strengthen evidential value, for instance whether studies were preregistered (notably because it constrains degrees of freedom in how data were handled and because severe testing bring more evidence for a hypothesis; Mayo & Spanos, 2006). Obviously, Kawakami et al. (2008) could not use this way to increase evidential value insofar as it was not used in social psychology at the time. Finally, a tool that can be used to assess evidential value, even when only few studies are available is the Bayes factor. Such index indicates how much we should update our belief that a hypothesis is true (usually compared to a null hypothesis) after observing empirical data (i.e., one or several studies; Jeffreys, 1961; Wagenmakers, 2007). Bayes factors thus make it possible to determine the level of evidence for a hypothesis in a study or a set of studies. Importantly, instead of simply looking at whether a replication study produces a significant *p* value or not, such Bayes factors quantify how much one learns from that replication.

In order to assess evidential value in favor of the hypothesis of a math approach avoidance training effect on implicit identification, one can therefore study either the aforementioned information or compute a Bayes factor based on Kawakami et al.'s (2008) data. For instance, one can observe that the critical *p* values were 0.060 and 0.048 (respectively for Studies 1 and 2) and that effect size estimations were rather imprecise (notably due to very small and small sample sizes respectively), with confidence intervals including or being close to a null effect size (see Fig. 4). Those two observations might suggest only weak evidence at this stage. In line with this conclusion, the Bayes factor we computed for these two critical tests reveals what is usually coined "moderate" evidence for this effect[1,2] ($BF_{10} = 7.36$; Jeffreys, 1961).

4. Increasing the evidence for the math approach training

Because we think Kawakami et al.'s (2008) paper offers an important leverage to reduce the gender gap in science regarding Nosek et al. (2009)'s results, but suffers from a quite modest evidential value, conducting a close replication is useful as it could strengthen the level of evidence for this effect. A successful replication would also support the idea of a generalizable effect, notably by using a different task, different stimuli, and a different population. In doing so, we will make sure to increase the sample size and to preregister the study, as this would increase the evidential value of the replication if the effect truly exists. In addition, we will adopt a different approach and avoidance training task (Rougier et al., 2018).

In their original experiments, Kawakami et al. (2008) adopted the Joystick task (without visual feedback) as an implementation of approach and avoidance. This task relies theoretically on the meaning associated with arm movements (for a seminal work in this area, see Cacioppo et al., 1993; Chen & Bargh, 1999): Arm flexion being associated with approach (bringing something toward the self) and arm extension being associated with avoidance (pushing something away). Other studies, however, have shown that arm flexion could be associated with avoidance (e.g., withdrawing one's hand from something) and arm extension with approach (e.g., approaching one's hand from something; Paladino & Castelli, 2008; Seibt et al., 2008). This suggests that those mappings are not hardwired at the motor level, but are open to cognitive interpretation (Seibt et al., 2008). Consistent with this idea, at least when studying approach/avoidance as a measure (to demonstrate that people are, for instance, faster to approach positive stimuli and avoid negative stimuli than to perform the reverse actions), effects seem more robust when approach/avoidance operationalizations rely

[1] Levels of evidence a Bayes factor can provide usually encompass "anecdotal", "moderate", "strong", "very strong", or "extreme evidence" for either a null hypothesis or a specific alternative (Jeffreys, 1961).

[2] Note that our computation of the Bayes factor differs from the recommendation of Field et al. (2019) regarding the assessment of evidential value for the purpose of selecting replication candidate. Indeed, we decided to use all the information available regarding the math approach training effect, resulting in the inclusion of Kawakami et al.'s (2008, Experiment 1) marginally significant math approach training effect (i.e., $p = .060$). As a consequence, evidence in favor of the math approach training might be overestimated. When this marginal effect is not included, the evidential value for the math approach training effect is only "anecdotal" ($BF_{10} = 1.70$; Jeffreys, 1961).

on movements of the whole self (where moving closer and away almost always means approaching and avoiding respectively) instead of arm movements (Krieglmeyer & Deutsch, 2010). This difference with movement of the whole self is even larger for arm movements operationalizations which do not use visual feedback in addition to arm movements (Krieglmeyer & Deutsch, 2010).

For these reasons, even though the current work studies approach/avoidance as a training and not as a measure, we decided to rely on an approach/avoidance operationalization that simulates visually movement of the whole self: the Visual Approach/Avoidance by the Self Task (VAAST; Rougier et al., 2018). In this task, participants receive instructions to approach and avoid categories of stimuli (e.g., "approach positive words and avoid negative ones") appearing in a virtual environment (e.g., a street), to do so, they have to press either an "approach" or "avoid" response key. Pressing the "approach" response key triggers an animation on the screen corresponding to an approach visual flow: The point of view within the environment gets further and the stimulus in front of the point of view gets bigger; pressing the "avoidance" response key triggers an avoidance visual flow.

By adopting the VAAST rather than the Joystick task for an approach/avoidance training, we adopt a task that emphasizes visual feedback in approach and avoidance movements rather than motor activation. In both procedures, participants receive the instructions to approach and to avoid different classes of stimuli. However, the Joystick task relies (theoretically) on the movement the participants perform while the VAAST focuses on the approach/avoidance visual feedback participants' responses entail. Because the latter are unambiguously mapped to either an approach or avoidance movement (Rougier et al., 2018), we think it is a safer implementation of approach and avoidance for an approach and avoidance training.[3] Another possible added value could be that, even though it will not be the case in the current studies conducted in the lab, the VAAST does not necessitate any particular material (like a joystick) and can easily be applied to online experiments (Aubé et al., 2019). This could be useful to increase sample sizes and the kind of targeted populations. Using the VAAST in order to replicate Kawakami et al.'s (2008) study, however, is viable only under the assumption that the VAAST can produce results at least as large as the Joystick task when both tasks are used as a training. We test this assumption in a pilot experiment.

5. Overview of the current research

In this manuscript, we conducted a preregistered replication of Kawakami et al.'s (2008) by 1) substantially increasing the sample size and 2) using the VAAST as an implementation of approach and avoidance (Rougier et al., 2018). We focused on the approach and avoidance training effect on identification IAT because Roland et al. (2018) showed that this measure is a good predictor of academic persistence. In a pilot experiment, we first tested whether the VAAST could be used to form indirect evaluations (i.e., on an identification IAT) and we checked whether it performed at least as well as the Joystick task, as implemented in Kawakami et al. (2008). This comparison was relevant, because in the event that we did not replicate Kawakami et al.'s (2008) results, doubts could be raised that it was due to our use of a different training task. Then, we replicated Kawakami et al.'s (2008) experiment with the VAAST, comparing a math approach training and a math avoidance training among women with low identification with math (see the replication recipe in Appendix A; Brandt et al., 2014). Finally, we conducted a meta-analysis including our replication as well as Kawakami et al.'s (2008) results. We report all measures, manipulations, and exclusions. Material, data, and analysis scripts can be found at osf.io/pcea3/.

6. Pilot experiment

The goal of the Pilot Experiment was twofold. First, we wanted to test whether a VAAST-based training could form a specific evaluation (on an identification IAT). Second, we wanted to make sure that any difference between the VAAST-based training and the Joystick task training used by Kawakami et al. (2008) would be negligible. If, for any reason, the Joystick task training effect was larger in this pilot experiment, it would have raised concerns regarding our goal to replicate Kawakami et al.'s (2008) work using the VAAST.

6.1. Method

6.1.1. Participant, design, and sensitivity analysis

One hundred and fifty (M_{age} = 20.10, SD_{age} = 2.48, 127 women) psychology students from a French university took part in this study in exchange for course credit. We relied on a 2 (training: Luupite approach vs. Niffite approach) by 2 (training task: Joystick task vs. VAAST) between-participant design.

The published literature suggests that the expected effect of approach and avoidance training is large (d_s = 1.04; Van Dessel et al., 2017). With the sample size of 140 participants we preregistered, we had enough power to detect the training effect (i.e., $1 - \beta > 0.99$). Moreover, we had enough power to conduct a negligibility test for the difference between the Joystick task and the VAAST (i.e., $1 - \beta = 0.80$).

Remember that we do not expect large differences in the effectiveness of the Joystick task and the VAAST. Negligibility tests allow considering an effect as negligible if the estimated effect size is smaller than an a priori limit. In other words, to ensure that the difference of effectiveness between the Joystick task and the VAAST was not large, we set an effect size interval in which the estimated effect of our study was likely to fall, if no effect existed in the population (e.g., 80% of the time). If the estimated effect in our study was outside this interval, we could not consider the difference between the Joystick and the VAAST negligible, even if the difference between the two was not significant (this results would be considered as non-informative). This analytic strategy is used to address pitfalls regarding interpretation of non-significant effect as null. In our preregistration, in order to satisfy power requirements suggested by Cohen (1988), we used an effect size of $\eta_p^2 = 0.013$ as our threshold (for an introduction on the concept of smallest effect size of interest, see Lakens, 2017).

6.1.2. Procedure and material

When arriving at the lab, participants were led to an individual cubicle. The experimenter then explained that the experiment would be about movement and response time and that they would engage in multiple tasks for a total time of 25 min. All the instructions were displayed on a computer screen. The experimenter also informed participants that they might have to use experimental material during the experiment and that this material would be located behind the computer screen. The experimenter was blind to participants' condition. Participants were seated at approximately 70 cm from the 24-in. screen. The experiment was programmed using *E*-prime©.

6.1.3. Approach and avoidance instructions

Two lists of words (i.e., the Luupite and the Niffite list; see Gregg et al., 2006) were presented to the participants. Every word in the Luupite list ends with -lup and had two consecutive vowels (e.g., maasolup, tuuralup) and every word in the Niffite list ends with -nif and

[3] By safer, we mean that it is less likely with the VAAST implementation that the ambiguity of approach and avoidance implementation interferes with the training. Note that we consider this interference nonsystematic and more likely to occur only in specific condition (e.g., when it is easier for the participant to perform a specific movement compared to another for a target category; for similar reasoning, see Krieglmeyer & Deutsch, 2010). That explains why we did not formulate the hypothesis that the VAAST should necessarily produce a larger effect than the Joystick task in the Pilot Study (see our pre-registration).

had two consecutive consonants (e.g., cellanif, otrannif). An exhaustive list of the stimuli can be found in Appendix A.

Participants' task for the first part of the experiment was to approach the words of one list and to avoid the words of the other. Half of the participants was randomly assigned to approach words from the Luupite list and avoid the one from the Niffite list, the other half had to do the opposite.

6.1.4. Approach and avoidance training

After receiving the instructions, participants performed the approach and avoidance training. They were randomly assigned either to the Joystick task training or the VAAST training.

6.1.5. Joystick task training

Participants performed an approach/avoidance task identical to the one used by Kawakami et al. (2008) but for the stimuli used. To approach the target word, participants pulled the joystick toward themselves (as if they were grabbing the word toward themselves) and to avoid the target word, they pushed the joystick away from themselves (as if they were pushing the word away).

For each trial, participants saw a fixation cross for a random duration (between 650 ms and 1200 ms) followed by a target word at the center of the screen. When the target word appeared, participants either pulled the joystick toward themselves or pushed away the joystick from themselves. In case of incorrect response, the "Error!" message appeared in a gray box and participants then had to start the trial over. Participants started by a practice block which consisted of 20 trials and then, completed the main block of 200 trials; one half of the trials was related to approach and the other one was related to avoidance. The training phase lasted approximately 10 min.

6.1.6. VAAST training

Participants in the VAAST training condition were informed that they would have to perform approach and avoidance movements within a virtual environment on the computer and that they would have to use the keyboard arrow keys of the directional keypad to do so (see Fig. 1). Participants pressed the up arrow when they had to approach the target word and pressed the down arrow when they had to avoid it.

Each trial started with a fixation cross for a random duration (between 650 ms and 1200 ms) followed by a target word at the center of the screen. When the word was presented, participants pressed the key corresponding to the movement they had to perform. When participants pressed the correct key, the screen simulated an approach/avoidance movement: The stimulus size, as well as the point of view within the environment, simulated a movement (see Rougier et al., 2018). Participants had to press the correct response key twice to complete a trial, each key press triggering an animation, as if they were walking two steps forward/backward within the environment. If participants made an error, as in the Joystick task condition, a gray alert box appeared on the screen informing them so and they had to start the trial over again. Participants started by a 20-trial practice block and then moved to a 200-trial main block; one half of the trials was related to approach and the other one was related to avoidance. The approach/avoidance phase lasted approximately 10 min.

6.1.7. Identification IAT

Immediately after the approach and avoidance training phase, participants completed an identification IAT (see Nosek et al., 2002). In this task, participants categorized words from the "Luupite" and "Niffite" lists as well as words related to the self (e.g., I, mine) and others (e.g., they, theirs). An exhaustive list of the stimuli can be found in Appendix A. To complete this task, participants used a *Chronos* response box.

This IAT was identical to the one adopted by Kawakami et al. (2008) but for the stimuli used which were related to Luupite and Niffite. We collected response latencies for the two 72-trial critical blocks. In one of the critical blocks, participants categorized Luupite and self-related words with a single response key (e.g., the E key) and Niffite and other-related words with another single response key (e.g., the I key). In the other critical block, participants categorized Niffite and self-related words with a single response key and Luupite and other-related words with another single response key. Order of these blocks was counterbalanced across participants.

After this task, participants were asked for their demographics. They

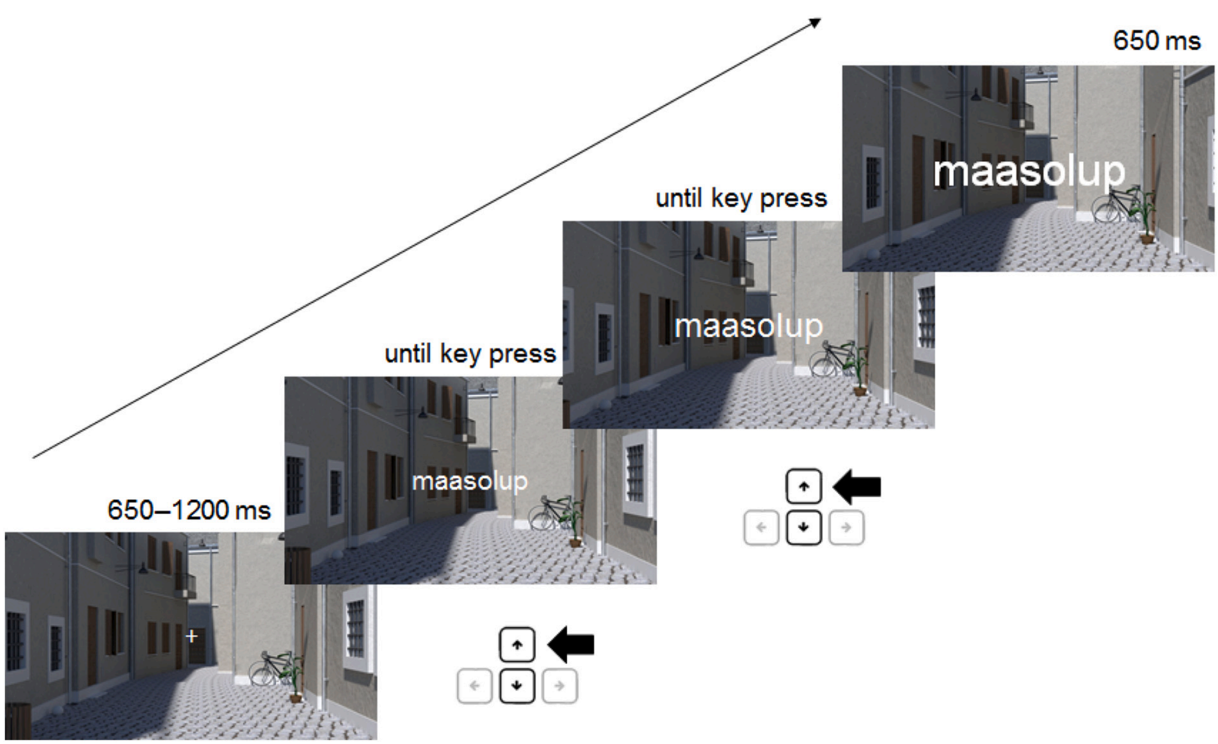

Fig. 1. Time course of a VAAST trial in the Pilot Experiment. Participant has to approach Luupite words. Black arrows represent a key press.

were then thanked and debriefed by the experimenter.

6.1.8. Analysis and discussion

We used the same IAT score calculation as Kawakami et al. (2008): We excluded incorrect trials (7.25% of the trials) and we recoded response times (RT) under 300 ms and above 2000 ms to respectively 300 ms and 2000 ms (0.95% of the remaining trials). Then, we log-transformed RT and computed the difference between critical blocks so that a positive score indicates that the participant was faster in the "self-Niffite" than in the "self-Luupite" block. We used this score as dependent variable in our analyses.

To test the hypothesis that an approach training toward a group of stimuli led to faster RT in the block where these stimuli shared the same response key with "self", we conducted a between-participant ANOVA.[4] Between-participant predictors were the approached target (Luupite approach vs. Niffite approach) and the task used during the training (Joystick task vs. VAAST). We excluded one statistical outlier of this analysis because of a gap on their Cook's D (Judd et al., 2017).[5]

As predicted, this analysis revealed that the identification IAT difference scores were significantly lower when participants approached Luupite words than when they approached Niffite words, $t(145) = 6.14$, $p < .001$, $CI_{95\%}[0.083; 0.163]$, $\eta_p^2 = 0.201$ (see Fig. 2). This effect indicates that participants were significantly faster in the block where the self and the list of words they approached were associated with the same response key. Additionally, the difference between the two tasks, descriptively in the direction of a superiority of the VAAST, was not significant, $t(145) = 1.15$, $p = .252$, $CI_{95\%}[-0.33; 0.125]$, $\eta_p^2 = 0.009$. Simple slope analysis revealed large movement effects for both the Joystick task and the VAAST, respectively, $t(145) = 3.33$, $p = .001$, $CI_{95\%}[0.043; 0.157]$, $d_s = 0.85$, and, $t(145) = 5.50$, $p < .001$, $CI_{95\%}[0.091; 0.201]$, $d_s = 1.18$. The main effect of the task used during the approach training was not significant, $t(145) = 1.75$, $p = .083$, $CI_{95\%}[-0.33; 0.125]$, $\eta_p^2 = 0.020$.

To avoid relying on a null effect, we preregistered that the difference between the two tasks could be considered as negligible. Wellek (2010, pp. 278–284) provides a procedure to conduct negligibility tests of interactions. In this procedure, one needs to set an a priori limit that the estimated effect size of the interaction is not supposed to exceed for the effect to be considered negligible. The preregistered value for this limit was $\eta_p^2 = 0.013$. Because our estimate effect size was $\eta_p^2 = 0.009$, the difference of effect between the tasks, in addition to not being significant, is negligible.

These results indicate that approaching and avoiding novel groups of stimuli influence their indirect evaluation: Participants are faster during the identification IAT when the self category and the approached list are associated to the same response key compared to when the self category and the avoided list are associated to the same response key. Given that the Joystick task effect was not significantly larger than the VAAST effect (and in fact descriptively smaller), because the VAAST comes with no ambiguity in response mapping, and because using a different task could increase generalizability, we used this task to conduct our Kawakami et al.'s (2008) replication.

7. Kawakami et al.'s replication

With this experiment, we aimed at replicating Kawakami et al.'s (2008) work on the effect of a math-approach training on an identification IAT among women with low identification with math. We expected participants' indirect evaluations of math to change after the training. Participants who approached math-related stimuli should be faster in the IAT block where math and self categories shared the same response key, compared to participants who avoided math-related stimuli. Note that even if we were primarily interested in indirect evaluation change, we also included (at the end of the experiment) a direct evaluation of math identification for exploratory purposes.

7.1. Method

7.1.1. Participant, design, and power analysis

Two hundred and three participants[6] ($M_{age} = 20.91$, $SD_{age} = 3.37$) were recruited for this experiment. Participants took part in the experiment in exchange for either class credit or 10€. We relied on a 2 (training: "math approach" vs. "math avoidance") between-participant design.

Note that even if the original and replication context (respectively, Canada and France) are not identical, they are similar on several points. First, both Canada and France are WEIRD countries, that is Western, Educated, Industrialized, Rich, and Democratic (Henrich et al., 2010). Moreover, in terms of gender gap in math achievement, OECD reports show quite similar numbers for Canada and France (OECD, 2015). These numbers indicate, for example, that boys outperform girls in problem solving in both countries, but also that the gender gap among top achievers in both Canada and in France is larger than the OECD average. We consider the context of replication as well as the sample used for the replication similar to the original Kawakami et al.'s (2008) paper.

Regarding statistical power, Funder et al. (2014) recommend adopting a more demanding threshold than the usual 80% recommended by Cohen (1988) when conducting a single study replication. Accordingly, we preregistered a 200 participants sample size for our replication as it allows us to reach a statistical power over 99% to detect an effect of Kawakami et al.'s (2008) meta-analytical effect size (i.e., $d_s = 0.64$). It is worth mentioning that one could argue that Kawakami et al.'s effect size might be inflated given the small sample sizes they adopted (Schönbrodt & Perugini, 2013). An alternative, safest, strategy could have been to ignore Kawakami et al.'s results and use the average effect size in social psychology instead ($d_s = 0.40$; Richard et al., 2003). A power analysis using this effect size reveals a 80% statistical power. Although this still complies with Cohen's (1988) recommendation, it can be considered suboptimal for a single study replication research (Funder et al., 2014).

7.1.2. Procedure and material

We conducted the replication following Kawakami et al.'s (2008) Experiment 1 procedure. In what follows, we explicitly mention the differences between the original experiment and the replication. A summary of the difference between the Kawakami et al.'s original experiment and the replication can be found in the replication recipe in Appendix A (Brandt et al., 2014).

Participants were pre-screened by answering a questionnaire either online or in person. We recruited only female participants who reported not liking math (i.e., less than four on a "I like math" seven-point Likert item). Participants who fitted our inclusion criteria were invited to take part in the actual experiment. We informed participants that they would have to engage in a series of tasks aiming at testing material related to school affinity.

One out of four different experimenters welcomed the participants in the lab and led them in an individual cubicle where participants received instructions on how to complete the study. Participants were randomly assigned to one of the two experimental conditions—math

[4] We report effect sizes and 95% Confidence Interval (CI) in our analyses. Effect sizes are reported following Lakens (2013) recommendation, that is η_p^2 in ANOVA design and Cohen's d_s for comparison between groups. 95% CI corresponds to the confidence interval for the parameter in the underlying model.

[5] Not removing the statistical outlier neither change the direction nor the significance of the results.

[6] Data of two participants were lost due to a power outage during the procedure.

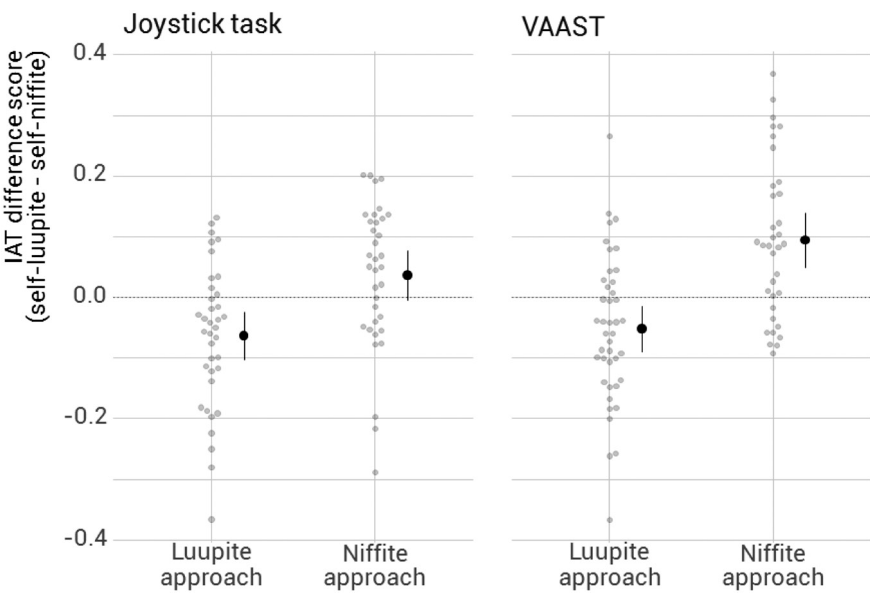

Fig. 2. IAT difference score according to the approach training condition (Luupite approach vs. Niffite approach) and the task used for the training (Joystick task vs. VAAST). IAT difference scores were computed by subtracting log-transformed RT for the block where "self" and "luupite" were associated with the same response key to log-transformed RT for the block where "self" and "Niffite" were associated with the same response key. Each point represents the score for a given participant. Participants above the dotted horizontal line are faster in the "Niffite-self" block, participants below are faster in the "Luupite-self" block. Error bars correspond to $CI_{95\%}$. RT = reaction times.

approach condition or math avoidance condition—and the experimenter was blind to this condition.

As in the Pilot Experiment, participants were first told that they would have to approach and avoid several stimuli in the first part of the experiment. Participants were randomly assigned to either the math approach condition or the math avoidance condition. Participants in the math approach condition were instructed to approach math-related pictures ($N = 24$) and to avoid arts-related ones ($N = 24$); participants in the math avoidance condition had the opposite instructions. A subset of the stimuli used for the approach and avoidance training can be found in Appendix A. Note that the pictures we used differ from Kawakami et al.'s (2008) because we could not access original material. Therefore, we sampled 48 stimuli from Wikimedia.

Unlike Kawakami et al.'s (2008) experiment which adopted a Joystick task implementation of approach and avoidance, the training in our replication was similar to the VAAST condition of the Pilot Experiment except for a few differences. Instead of words, participants approached or avoided pictures related to math or to arts. Instead of answering with keyboard directional buttons, participants had to press buttons on a *Chronos* response box. Approach and avoidance buttons were labeled with the "Approach" and "Avoid" labels. The total number of trials was 480 distributed in 10 blocks of 48 trials. One half of the trials was related to approach and the other one was related to avoidance. The training phase lasted approximately 25 min.

After the approach and avoidance training, participants completed an identification IAT with math and arts as target categories. The IAT procedure was the same as in the Pilot Experiment. An exhaustive list of the stimuli used for the IAT can be found in Appendix A.

A notable difference with Kawakami et al.'s (2008) original experiment is that we also decided to include a self-report measure assessing how much participants identify with math and arts. This measure is inspired by the Inclusion of Other in the Self scale (Aron et al., 1992), which was developed in the context of intergroup relationship and has already been used in the context of math identification (Necka et al., 2015). Participants were presented two circles and they could adjust how much the circles overlapped using the keyboard arrow keys. They were informed that one circle was representing themselves and the other was representing the concept of math (arts). Participants received the instruction to choose the overlap between the circles corresponding to how much they thought their sense of themselves was overlapping with the concept of math (arts). Participants answered for both math and arts concepts and the order was randomized across participants.

After this task, participants were asked for demographic information. They were then thanked and debriefed by the experimenter.

7.2. Analysis

7.2.1. Identification IAT

As in the Pilot Experiment, we used the same IAT score calculation as Kawakami et al. (2008): We excluded incorrect trials (6.23% of the trials) and we recoded RT under 300 ms and above 2000 ms (0.79% of the remaining trials) to respectively 300 ms and 2000 ms. Then, we log-transformed RT and computed the difference between test trials on critical blocks so a positive score indicated that the participant was faster in the "self-math" than in the "self-arts" block. We used this score as a dependent variable.

We conducted a *t*-test comparing participants who approached math-related material and those who avoided math-related material. We found a significant difference between the two conditions indicating that participants in the approach math condition had higher scores ($M = -0.051$, $SD = 0.117$, $n = 104$) than participants in the avoid math condition ($M = -0.084$, $SD = 0.106$, $n = 97$), $t(199) = 2.10$, $p = .037$, $CI_{95\%}[0.002; 0.064]$, $d_s = 0.30$. This analysis revealed a small effect in the direction of our prediction (see Fig. 3).

7.2.2. Self-overlap measure

We analyzed data from the self-overlap measure according to a mixed effect ANOVA with one between-participant predictor (i.e., training) and a within-participant predictor (i.e., target participants evaluated). We coded participants responses so that a score of 0 indicates that the circles representing the self and the concept do not overlap and so a score of 100 indicates the two circles totally overlap.

This analysis revealed a main effect of the target, indicating that participants reported more self-concept overlap when the target was "arts" ($M = 62.63$, $SD = 25.46$, $n = 201$) than when the target was "math" ($M = 20.82$, $SD = 20.05$, $n = 201$), $t(199) = 16.26$, $p < .001$, $CI_{95\%}[36.83; 46.99]$, $\eta_p^2 = 0.569$. The main effect of the training was not significant, participants in the "math approach" condition ($M = 40.87$, $SD = 13.30$, $n = 104$) did not report higher overlap than participants in the "math avoidance" condition ($M = 42.64$, $SD = 14.38$, $n = 97$), $t(199) = 0.91$, $p = .364$, $CI_{95\%}$ $[-5.63; 2.07]$, $\eta_p^2 < 0.001$. The interaction between the training condition and the target participants evaluated was not significant, $t(199) = 1.10$, $p = .271$, $CI_{95\%}[-15.86; 4.48]$, $\eta_p^2 = 0.002$: The difference between how participants overlapped with math and arts concept was not

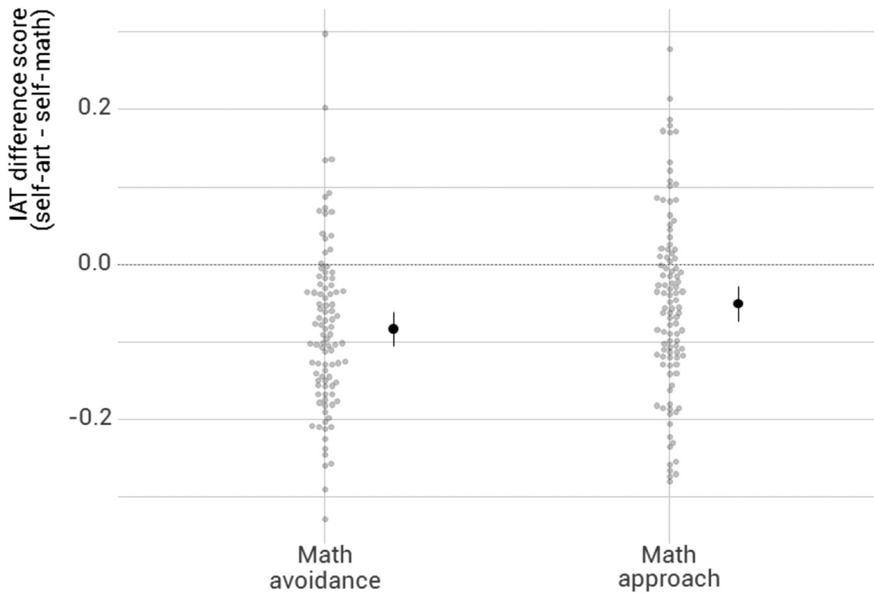

Fig. 3. Identification IAT difference score according to the experimental condition (Math avoidance vs. Math approach). IAT difference scores were computed by subtracting log-transformed RT for the block where “self” and “arts” were associated with the same response key to log-transformed RT for the block where “self” and “math” were associated with the same response key. Each point represents a participant. Participants above the dotted horizontal line are faster in the “math-self” block, participants below are faster in the “arts-self” block. Error bars corresponds to $CI_{95\%}$. RT = response times.

significantly reduced after a math approach training.

8. Complementary analyses

To better evaluate the contribution of the present experiment regarding Kawakami et al.'s (2008) original data on the effect of math approach training on math identification, we computed a Bayes factor using the Savage-Dickey density-ratio method (Wagenmakers et al., 2010). A Savage-Dickey Bayes factor reveals how much our belief in a hypothesis changes after collecting data. This analysis revealed that the hypothesis of a null effect of the math approach and avoidance training is now 2.57 less likely. By computing a meta-analytic Bayes factor, we estimated that the current level of evidence for an effect of a math approach training compared to no effect is now “strong” ($BF_{10} = 18.77$; Jeffreys, 1961; Morey & Rouder, 2018).

Moreover, to summarize the current evidence for a math approach and avoidance training effect on math implicit identification among women with low identification with math, we conducted a meta-analysis. We included Experiments 1 and 2 of Kawakami et al. (2008), as well as our replication for a total of 3 effects (i.e., $k = 3$) and 281 participants. We conducted a random-effects meta-analysis with the *metafor* package (REML estimation; Viechtbauer, 2010). The estimated effect size of the approach and avoidance training effect was $d_s = 0.40$, d_s $CI_{95\%}$ [0.14; 0.65] and there was no evidence for variation in the effect sizes between studies, $Q(2) = 1.74$, $p = .419$. This meta-analysis suggests a small-to-medium effect size for the effect of approaching math-related material (instead of avoiding it) on an identification IAT (see Fig. 4).[7]

[7] Given that our replication adopted a different approach and avoidance paradigm than the original work we replicated, readers might wonder why we did not investigated the difference in procedure effectiveness between the two paradigms. The reason is that we consider that the outcome the most plausible of such analysis would be an absence of significant difference. This is because the original Kawakami et al.'s (2008) estimation of the effect size is rather imprecise. Because of this, a difference between our replication and the original work would be significant only if the effect size of the replication is below an effect size of $d_s = 0.09$ or above an effect size of $d_s = 1.20$. In our replication, an effect as small as $d_s = 0.09$ could not be reliably estimated given the sample size of the replication and an effect size as large as $d_s = 1.20$ would be unrealistic (for a benchmark of effect sizes in the implicit evaluation change literature, see Lai et al., 2014).

9. General discussion

Reducing the gender gap related to math is still a major challenge in our society and implicit attitude change offers an interesting perspective on this issue (Nosek et al., 2009). Kawakami et al. (2008) observed that women who repeatedly approached math-related material self-identified more to math (as measured with an identification IAT; Nosek et al., 2002). This original experiment, however, suffered from a quite modest evidential value. With the present replication, we aimed at providing stronger evidence in favor of the math approach training effect, as well as increasing the generalizability of Kawakami et al.'s (2008) finding. We conducted this replication with a sample size large enough to detect small-to-medium effect size and we preregistered our hypotheses and analytic strategy. Using a VAAST-based procedure (Rougier et al., 2018), we replicated Kawakami et al.'s (2008) results, providing stronger evidence for an effect of math approach training (i.e., increasing the evidential value by a factor of 2.57). Building upon Kawakami et al. (2008) and endorsing today's research standards (see Vazire, 2018), this replication therefore contributes to provide experimental evidence that an approach and avoidance training can change identification. Beside the results on the identification IAT, we decided to explore the impact of the approach and avoidance training on a more direct measure of identification (i.e., the self-overlap measure). Several reasons might explain the absence of significant difference between participants who approached math and those who avoided math. First, the statistical noise of a single-item measure makes the comparison between the conditions on the self-overlap measure inherently less powerful than the comparison on the IAT. However, recent results in the literature suggest that one can acquire a new indirect evaluative response, without having an impact on a more direct one. Indeed, Van Dessel et al. (2016) obtained such pattern of results in an experiment focusing on the effect of approach and avoidance instructions. Still, at this stage, one can probably infer that approach and avoidance training might not be the easiest way to change explicit identification with math.

While Kawakami et al. (2008) relied on the idea that the mere association between a movement and a stimulus has an effect on implicit identification, a recent theoretical account has suggested inferential processes as a mechanism to explain the effect of approach and avoidance paradigm. According to Van Dessel et al. (2018) performing an action influences evaluations because participants infer a relation between the stimulus and evaluation due to the available information during the training. That is, when participants have to repeatedly

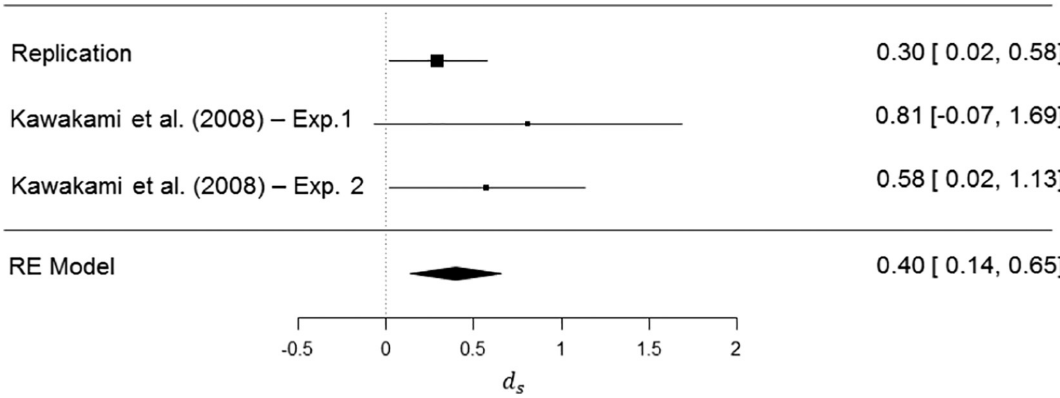

Fig. 4. Random-effect meta-analysis on the math approach and avoidance training effect on a math identification IAT. Effect sizes have been coded so that positive d_s indicates an increase in math identification after a math approach training.

perform an approach movement faced to a stimulus, the participants would infer an association between the stimulus and the approach concept. Given that participants know that they usually approach things that are relevant to them, the approached stimulus would be tied to the self. Critically for the inferential account, when participants can infer the same information out of instructions they would have from the training, they do acquire the same evaluative response (Van Dessel et al., 2015; Van Dessel et al., 2016).

One could thus argue that an actual training would not be the most efficient way to change implicit identification as instructions alone have an effect on indirect evaluations. Nevertheless, Van Dessel et al. (2015) failed at showing an effect of the instructions for stimuli for which people already had strong attitudes (i.e., ethnic groups). Moreover, Van Dessel et al. (2020) recently showed that, when it comes to real social groups (i.e., Turkish and Flemmish groups), actual training was an effective way to change implicit attitude while instruction was not. Our results are in line with the literature, suggesting that actual experience in interventions aiming at changing implicit identification possesses some critical features that are worth investigating when studying the change of strong evaluative responses. The current work is therefore an important contribution that confirms the robustness of such interventions.

Regarding the effectiveness of the approach avoidance training on implicit identification with math, the overall effect size estimated from Kawakami et al.'s (2008) experiments and the current replication corresponds to a small-to-medium effect. Despite the differences between Kawakami et al.'s (2008) original work and our replication in terms of sample (Canadian vs. French), time (2008 vs. 2017), paradigm (Joystick task vs. the VAAST) and material (the pictures used in the training task), we retrieved a math approach training effect in our replication. This suggests generalizability of the approach training as a way to change math implicit identification. Moreover, this effect size is in line with the literature interested in interventions aiming at changing indirect evaluation. Lai et al. (2014) conducted a large-scaled study where they compared the effectiveness of different interventions aiming at reducing implicit racial preference. The most effective interventions could change implicit preference with effect sizes ranging up to d_s = 0.49. In comparison, the estimated math approach and avoidance training had an effect size of d_s = 0.40. The effectiveness of the approach and avoidance training intervention is, thus, close to the most effective interventions included in Lai et al. (2014) comparison. This highlights the importance to consider approach and avoidance training when it comes to changing implicit identification.

One limitation to mention, however, as to do with the sample size we used in this replication. As we discussed in our power analysis section, we performed our power analysis by relying on the effect sizes observed in Kawakami et al. (2008)'s work, but (their) small sample sizes also come with unprecise effect sizes. Hence, our study can be seen as underpowered (i.e., 80% power instead of 90%, although this could be somewhat compensated by using a conservative two-tailed test) if one instead uses the average effect size in social psychology (Richard et al., 2003).

10. Conclusion

What can be concluded for approach and avoidance training when it comes to implicit identification change? With this high-powered close replication, we showed that approach and avoidance training can change implicit identification, successfully replicating Kawakami et al.'s (2008) results. To sum up, this replication project endorsing latest standards in terms of research practices (i.e., preregistration, open material, open data) increased the generalizability of Kawakami et al.'s (2008) finding by adopting a different sample and a different implementation of approach and avoidance (namely, the VAAST, Rougier et al., 2018 and a new material) and increased the level of evidence for an effect compared to a null effect by a factor of 2.57. Thus, we believe it offers more evidential value to the math approach training effect originally hypothesized by Kawakami et al.'s (2008).

Open practices

Materials (in French), data sets, analysis scripts, and preregistration forms can be accessed in the following OSF project: osf.io/pcea3/.

Funding

It was supported by Cédric Batailler PhD funding grant from Alpes Grenoble Innovation Research (AGIR) as well as a Pôle Grenoble Cognition and NeuroCoG IDEX UGA grant in the framework of the “Investissements d'avenir” program (ANR-15-IDEX-02).

Declaration of competing interest

The authors declare no conflict of interest.

The authors are grateful to Anaëlle Brion, Damien Foinant, and Caitlin Troussier-Thevenot for collecting the data on the replication study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2020.104059.

References

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science, 12*(5), 385–390. https://doi.org/10.1111/1467-9280.00371.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596.

Aubé, B., Rougier, M., Muller, D., Ric, F., & Yzerbyt, V. (2019). The online-VAAST: A short and online tool to measure spontaneous approach and avoidance tendencies. *Acta Psychologica, 201*, 102942. https://doi.org/10.1016/j.actpsy.2019.102942.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., … Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour, 2*(1), 6–10. https://doi.org/10.1038/s41562-017-0189-z.

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., … van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology, 50*(1), 217–224. https://doi.org/10.1016/j.jesp.2013.10.005.

Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes: II. Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology, 65*(1), 5–17. https://doi.org/10.1037/0022-3514.65.1.5.

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for Bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science, 2*(2), 115–144. https://doi.org/10.1177/2515245919847196.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin, 25*(2), 215–224. https://doi.org/10.1177/0146167299025002007.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology, 24*(1), 252–287. https://doi.org/10.1080/10463283.2014.892320.

Field, S. M., Hoekstra, R., Bringmann, L., & Van Ravenzwaaij, D. (2019). When and why to replicate: As easy as 1, 2, 3? *Collabra: Psychology, 5*(1), 46. https://doi.org/10.1525/collabra.218.

Funder, D. C., Levine, J. M., Mackie, D. M., Morf, C. C., Sansone, C., Vazire, S., & West, S. G. (2014). Improving the dependability of research in personality and social psychology: Recommendations for research and educational practice. *Personality and Social Psychology Review, 18*(1), 3–12. https://doi.org/10.1177/1088868313507536.

Giner-Sorolla, R. (2016). Approaching a fair deal for significance and other concerns. *Journal of Experimental Social Psychology, 65*, 1–6. https://doi.org/10.1016/j.jesp.2016.01.010.

Gonzales, P., Guzman, J. C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., & Williams, T. (2004). *Highlights from the trends in international mathematics and science study (TIMSS) 2003 [data set]*. https://doi.org/10.1037/e672832007-002.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109*(1), 3–25. https://doi.org/10.1037/0033-295X.109.1.3.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464.

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology, 90*(1), 1–20. https://doi.org/10.1037/0022-3514.90.1.1.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature, 466*(7302), 29. https://doi.org/10.1038/466029a.

Jeffreys, H. (1961). *Theory of probability*. Oxford, England: Oxford University Press.

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2017). *Data analysis: A model comparison approach to regression, ANOVA, and beyond* (3rd ed.). New York London: Routledge, Taylor & Francis Group.

Kawakami, K., Steele, J. R., Cifa, C., Phills, C. E., & Dovidio, J. F. (2008). Approaching math increases math = me and math = pleasant. *Journal of Experimental Social Psychology, 44*(3), 818–825. https://doi.org/10.1016/j.jesp.2007.07.009.

Krieglmeyer, R., & Deutsch, R. (2010). Comparing measures of approach–avoidance behaviour: The manikin task vs. two versions of the joystick task. *Cognition & Emotion, 24*(5), 810–828. https://doi.org/10.1080/02699930903047298.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J.-E. L., Joy-Gaba, J. A., … Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General, 143*(4), 1765–1785. https://doi.org/10.1037/a0036260.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00863.

Lakens, D. (2017). Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*(4), 355–362. https://doi.org/10.1177/1948550617697177.

Lakens, D., & Evers, E. R. K. (2014). Sailing from the seas of chaos into the corridor of stability: Practical recommendations to increase the informational value of studies. *Perspectives on Psychological Science, 9*(3), 278–292. https://doi.org/10.1177/1745691614528520.

Mayo, D. G., & Spanos, A. (2006). Severe testing as a basic concept in a Neyman–Pearson philosophy of induction. *The British Journal for the Philosophy of Science, 57*(2), 323–357. https://doi.org/10.1093/bjps/axl003.

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of bayes factors for common designs [manual]. https://CRAN.R-project.org/package=BayesFactor.

Necka, E. A., Sokolowski, H. M., & Lyons, I. M. (2015). The role of self-math overlap in understanding math anxiety and the relation between math anxiety and performance. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01543.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology, 83*(1), 44–59. https://doi.org/10.1037//0022-3514.83.1.44.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., … Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences, 106*(26), 10593–10597. https://doi.org/10.1073/pnas.0809921106.

OECD (2015). The ABC of gender equality in education. *OECD Publishing.*. https://doi.org/10.1787/9789264229945-en.

Paladino, M.-P., & Castelli, L. (2008). On the immediate consequences of intergroup categorization: Activation of approach and avoidance motor behavior toward ingroup and outgroup members. *Personality and Social Psychology Bulletin, 34*(6), 755–768. https://doi.org/10.1177/0146167208315155.

Richard, F. D., Bond, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7*(4), 331–363. https://doi.org/10.1037/1089-2680.7.4.331.

Roland, N., Mierop, A., Frenay, M., & Corneille, O. (2018). Field-identification IAT predicts students' academic persistence over and above theory of planned behavior constructs. *Frontline Learning Research, 5*(1), 19–30. https://doi.org/10.14786/flr.v6i1.327.

Rougier, M., Muller, D., Ric, F., Alexopoulos, T., Batailler, C., Smeding, A., & Aubé, B. (2018). A new look at sensorimotor aspects in approach/avoidance tendencies: The role of visual whole-body movement information. *Journal of Experimental Social Psychology, 76*, 42–53. https://doi.org/10.1016/j.jesp.2017.12.004.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009.

Seibt, B., Neumann, R., Nussinson, R., & Strack, F. (2008). Movement direction or change in distance? Self- and object-related approach–avoidance motions. *Journal of Experimental Social Psychology, 44*(3), 713–720. https://doi.org/10.1016/j.jesp.2007.04.013.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143*(2), 534–547. https://doi.org/10.1037/a0033242.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4–28. https://doi.org/10.1006/jesp.1998.1373.

Van Dessel, P., De Houwer, J., Gast, A., Roets, A., & Smith, C. T. (2020). On the effectiveness of approach-avoidance instructions and training for changing evaluations of social groups. *Journal of Personality and Social Psychology.*. https://doi.org/10.1037/pspa0000189.

Van Dessel, P., De Houwer, J., Gast, A., Smith, C. T., & De Schryver, M. (2016). Instructing implicit processes: When instructions to approach or avoid influence implicit but not explicit evaluation. *Journal of Experimental Social Psychology, 63*, 1–9. https://doi.org/10.1016/j.jesp.2015.11.002.

Van Dessel, P., De Houwer, J., Gast, A., & Tucker Smith, C. (2015). Instruction-based approach-avoidance effects: Changing stimulus evaluation via the mere instruction to approach or avoid stimuli. *Experimental Psychology, 62*(3), 161–169. https://doi.org/10.1027/1618-3169/a000282.

Van Dessel, P., Gawronski, B., Smith, C. T., & De Houwer, J. (2017). Mechanisms underlying approach-avoidance instruction effects on implicit evaluation: Results of a preregistered adversarial collaboration. *Journal of Experimental Social Psychology, 69*, 23–32. https://doi.org/10.1016/j.jesp.2016.10.004.

Van Dessel, P., Hughes, S., & De Houwer, J. (2018). How do actions influence attitudes? An inferential account of the impact of action performance on stimulus evaluation. *Personality and Social Psychology Review, 23*(3), https://doi.org/10.1177/1088868318795730.

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science, 13*(4), 411–417. https://doi.org/10.1177/1745691617751884.

Viechtbauer, W. (2010). Conducting meta-analyses in *R* with the metafor package. *Journal of Statistical Software, 36*(3), https://doi.org/10.18637/jss.v036.i03.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14*(5), 779–804. https://doi.org/10.3758/BF03194105.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the savage–dickey method. *Cognitive Psychology, 60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001.

Wang, M.-T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review, 29*(1), 119–140. https://doi.org/10.1007/s10648-015-9355-x.

Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority* (2nd ed.). Boca Raton: CRC Press.