The effectiveness of the "But-you-are-free" technique: Meta-analysis and re-examination of the technique

Adrien Alejandro Fillon[1], Lionel Souchet[2], Alexandre Pascual[3], and Fabien Girandola[2]

[1]ERA Chair in Science and Innovation Policy & Studies, University of Cyprus
[2]Department of Social Psychology, Aix-Marseille University
[3]Department of Social Psychology, University of Bordeaux

The "But you are free..." (BYAF) technique is a technique to increase compliance (for example, to give spare change for the bus), by adding the words "But you are free to accept or refuse" to the request. In this pre-registered meta-analysis, we examine the effect of the BYAF technique in 52 experiments ($N = 19528$). An analysis of 74 effect sizes showed a medium effect ($g = 0.44$, 95% confidence intervals (CI) [0.36, 0.51]) for the BYAF technique. A moderator analysis found a stronger effect for face-to-face interactivity over other types of interactivities. All the other moderators we used were not statistically significant. We did not find any differences between articles published before and after Carpenter's (2013) meta-analysis. An examination of risk of bias showed that only seven studies were of "low risk", and a meta-analysis of these studies showed no effect of the BYAF ($g = 0.11$, 95% CI [-0.18, 0.40]) We also found that most recent studies on the subject are too low-powered to detect the effect found by Carpenter (2013), and the reproducibility rates were critically low (R-index = 9.77%, Z-curve expected discovery rate = 6%). We propose some improvements to the design and experiments to ensure the effects found in the literature exist and are replicable. All materials are available on https://osf.io/8eqa5/

Keywords: But-you-are-free, Commitment, Compliance, Meta-analysis

Introduction

The But-You-Are-Free (BYAF) technique is a commitment technique invented and used by Guéguen and Pascual (2000). This technique consists of an addition of the words "but you are free" during a request to enhance the acceptance of the request. The BYAF technique is one of many techniques (see Pratkanis, 2007 for a review) used as commitment techniques, based on the reactance theory (Brehm, 1966). Contrary to other techniques, the BYAF is easy to use – you only need to add one sentence to the request. For example, Guéguen and Pascual (2000) observed 10% of compliance rate with the request "Sorry Madam/Sir, would you have some coins to take the bus, please?" (Control condition), whereas 47.5% was obtained with "Sorry Madam/Sir, would you have some coins to take the bus, please? But you are free to accept or to refuse." (BYAF condition). The "BYAF" technique can be combined to other techniques such as the "foot in the door" technique to further increase compliance. Furthermore, this technique can be applied in many situations, such as face-to-face interaction, but also in indirect interaction, for example with the use of the internet (e.g., e-mail, (Pascual, 2002). But how does it work? The exact

wording (i.e., "but you are free") is not required to enhance compliance, as other wording "but obviously do not feel obliged" (Guéguen et al., 2013, p. 129) is as effective. The technique relies on the salience of the target's freedom in their decision-making process. The acknowledgement that one can say "no" leads to say "yes" more often, and to be more committed, as shown by the amount of money given in most of the studies (e.g., Guéguen and Pascual, 2000). As commitment theory (Kiesler, 1971; Kiesler and Sakumura, 1966) postulates, it is possible to manipulate the degree of commitment by manipulating the degree of perceived choice when performing the act. As such, the BYAF technique can be considered as a non-pressure manipulation used to enhance compliance.

Original study and its follow-up

In the original study (Guéguen and Pascual, 2000), researchers indicated in the subject section that four confederates, 2 men and 2 women on average age of 20-22 years old asked 40 men and 40 women chosen at random in the street. In the Procedure section, they indicated that the experiment was made in a mall. In the control condition, the confederates say "Sorry

Madam/Sir, would you have some coins to take the bus, please?" and in the BYAF condition "Sorry Madam/Sir, would you have some coins to take the bus, please? But you are free to accept or to refuse." The confederate then noted if the participant accepted to give money, and noted the amount given before giving it back to the participant and debriefed him. In the result section, researchers indicated that 10% of subjects accepted the request in the control group, and 47.5% in the BYAF group, whereas the mean amount was 0.48$ in the control group and 1.04$ in the BYAF group (all differences were statistically significant at an α level of 0.05). Researchers indicated that this experiment shows the effectiveness of the BYAF technique to increase the probability of compliance, in saying yes to the request, and the implication of the subject, in giving a higher amount of money. In 2013, Carpenter conducted a meta-analysis of the BYAF technique with 42 studies published after the original described above. His goal was to summarize the effect size of this technique and show some probable mediators and moderators. Indeed, researchers wanted to show if face-to-face interaction was important to the BYAF technique, and if the type of choice (prosocial, offer, or selfish) and the time of the request (immediate or delayed) influenced the BYAF effect. Also, as the first research are based on a monetary request, it was important to assess that BYAF works in another context, such as a signature for a petition. The meta-analysis showed that the sample-size weighted correlation between the presence and absence of the BYAF technique and the proportion of those who complied with the request was $r = .13$ (i.e., $d = 0.26$), which is, according to the author, a moderate-sized increase in effectiveness with the use of the BYAF technique. It is typically considered a small to medium effect size (Sawilowsky, 2009). The sampling error explained 22% of the variation in effect size. The confidence interval of the correlation was not reported. Carpenter identified several moderators. An immediate request led to an $r = .18$ (i.e., $d = 0.37$), and a delayed request to an $r = .07$ (i.e., $d = 0.14$), which showed the importance to position the BYAF technique close to the targeted request. Prosocial requests were as likely to work ($r = .16$, $d = 0.32$) as selfish requests ($r = .16$, $d = 0.32$). Concerning the analysis of publication bias, Carpenter correlated the sample sizes and effect sizes and found an $r = -.30$. This result means that there is the possibility that, as the sample size increases, the effect decreases, potentially to a null effect. This result suggests that publication bias is present and that the effect size estimate is inflated. Thus, the actual effect size might be small. Also, researchers used the trim-and-fill technique (Duval and Tweedie, 2000) but did not provide

the plot associated. The trim-and-fill technique leads to a reduction of the effect size by .04 (from an $r = .13$ to an $r = .09$, $d = 0.18$). Some meta-analysts indicated that the Trim and fill technique performs poorly in the presence of substantial between-study heterogeneity (J. Higgins, Chandler, et al., 2022). Finally, as Carpenter pointed out, nearly all the experiments were conducted either by Guéguen or Pascual (see Table 1), but they found the strongest and the smallest effect sizes for the technique. One major problem of the Carpenter (2013) meta-analysis is some studies were flagged as of risk of having fabricated data (Brown, 2020). The flagged studies have the strongest effect sizes found (Odds Ratio for Dufourcq-Brana et al., 2006 $OR = 6.57$; Guéguen and Pascual, 2000 $OR = 8.14$; Pascual and Guéguen, 2002 $OR = 6$), thus, eliminating these results from our analysis might show a null effect of the BYAF technique. Also, it is possible that research on this subject improves over time, with larger sample sizes, and stronger methods, leading to convergence to the "true" effect size of the BYAF technique on compliance. In most cases in psychology, the original effect sizes are inflated (Schäfer and Schwarz, 2019). This is the reason why we conduct a novel preregistered and open meta-analysis on the BYAF technique over compliance, with a look at the inconsistencies we can find between our analysis and the one from Carpenter (2013).

Moderators

We want to investigate the moderators that can influence the effect of the BYAF technique. The research on the subject shows that the moderators that can influence the technique are the type of request (pro-social vs. selfish), the temporality (immediate or delayed), the gender of the subject and of the confederate (man vs. woman), the culture (individualistic vs. collectivistic), the interactivity (face-to-face vs. indirect), and the type of freedom evocation ("but-you-are-free" vs. other). We also want to test if there are substantial differences between the effect sizes found before and after the Carpenter (2013) meta-analysis.

Type of request

As Carpenter (2013) pointed out, the effectiveness of the BYAF technique might rely upon the type of request. For Carpenter and Boster (2009), the compliance-gaining techniques work better for pro-social benefits, like giving to a charity, rather than for selfish reasons, like giving to take the bus. Nonetheless, Carpenter (2013) found no difference in compliance rate for the pro-social and selfish types of requests. We seek to redo the analysis with the same hypothesis, given that the larger number of studies involved could give a better

estimate of the effect size, and possibly could detect a moderator effect of the type of request. In doing so, our hypothesis is the same as in Carpenter's (2013) meta-analysis: the compliance rate will be higher for the pro-social type of request than for the selfish type of request.

Temporality

Temporality was called "immediate or delayed" in Carpenter's (2013) analysis. Indeed, depending on the studies, the researchers can look at whether the participant complied with the request immediately after using the technique (e.g., when they asked for money, the original technique), or after a certain amount of time (e.g., by sending an email and then testing at whether the participant had made a purchase, Grassini et al., 2012). We seek to replicate the effect of temporality found in the Carpenter's meta-analysis. Researchers found that the compliance rate was lower when the confederate was absent (delayed condition) than present (immediate condition). Two reasons are possible: an easier reactance involved with the absence of the confederate, or the wanting to have a better self-representation when the researcher is present. We seek to redo the analysis with the addition of new studies to find that the immediate use of the BYAF technique is more effective than the delayed use.

Subject gender

Studies seem to indicate that men are less compliant than women (Grosch and Rau, 2016). For example, one study found that men cheat more than women (Fischbacher and Föllmi-Heusi, 2013). Grosch and Rau (2016) indicated that this difference can be explained by the cultural roles of men and women, as women are seen as more pro-social than men. Thus, we think that Female participants will comply more to the request in the BYAF condition than men.

Confederate gender

Many experiments have shown that confederate gender influences the compliance rate. For example, Vaughn et al. (2009) have only found an effect of compliance when the confederate was a woman. Long et al. (1996) found that women were more helped than men. On the contrary, Dolinska and Dolinski (2006) found that both sexes have a better chance to find compliers when confederate sex matches the participant sex. This difference can be explained by cultural variation. Since most of the experiments were conducted in France, we think that the BYAF technique will be more effective if the confederate is a woman. Indeed, we hypothesize that participants will comply more to women confederate than to men confederate in the BYAF condition.

Culture

In pro-social culture such as in China, one could expect more compliance than in a more individualistic culture such as in France (see Hamamura et al., 2018). There are at least three reasons for this hypothesis. On general, the theory of commitment is more effective for individualistic than for collective culture (Kim and Sherman, 2007), because people in individualistic culture have a more internal locus of control (Channouf, 1990; Desrumaux, 1996), and people are more easily reactant (Jonas et al., 2009). Thus, the BYAF technique which reduces reactance should work better for people in an individualistic culture. Indeed, Pascual et al. (2012) showed that the BYAF technique induces more compliance in individualistic countries (i.e., France, Romania) than collectivistic countries (i.e., Ivory Coast, China, and Russia). According to Triandis (1989), individualist cultures include Northern and Western Europe as well as North America, whereas collectivist cultures would be characteristic of Asia, Africa, and South America. Participants from an individualistic country would comply more with the BYAF technique than participants from a collectivistic country.

Interactivity

If the BYAF technique has a different effect depending on the gender of the participant, or/and the gender of the confederate, it implies that this difference is within a "face-to-face" interaction. Furthermore, the difference between temporality (immediate or delayed), implies a difference between a "face-to-face" interaction and more distal interactions. We believe that participants are more engaged when the interaction is in "face-to-face" rather than in a more indirect interaction, via email, phone call, or internet.

Type of freedom evocation

The BYAF technique is an induction in a sentence (typically "but you are free to accept or to refuse") and induce a feeling of freeness making the recipient more willing to accept the demand, or to comply. Other evocations include propositions such as "do not feel obliged", "do as you wish", or "feel free to refuse". There are possibilities that some evocations are better than others to induce compliance. Indeed, the proposition "but you are free to refuse" is the most salient, leading to the best understanding by the recipient that he/she is free to accept or not. It should have a stronger effect on compliance than the other possibilities of evoking freedom.

Before and After Carpenter's analysis

Garmendia et al. (2019) have shown that 46% of meta-analyses have their conclusions altered by false data, with fraudulent/plagiarized studies, or errors. As we previously showed, Carpenter analysis has this problem. Original effect sizes are inflated (Schäfer and Schwarz, 2019) and we tend to think that most recent research is of better quality than before the crisis in social science (Motyl et al., 2017). In Carpenter's (2013) meta-analysis, the use of the Trim-and-Fill method reduced the effect size found close to the null, we hypothesize that the effect will be lower after Carpenter's analysis than before.

Summary hypotheses

Main hypotheses

People tend to comply more with the "but you are free" technique than with direct asking.

Confirmatory hypotheses

The compliance rate will be higher for 1) the prosocial type of request than for the selfish type of request and 2) immediate asking than delayed.

Exploratory hypotheses

The compliance rate will be higher (a) for women than for men, (b) for women confederate than for men confederate, (c) from an individualistic country than from a collectivistic country, (d) in a "face to face" interaction than in other types of interaction, (e) with the exact proposition "but you are free" than the others types of evocation and (f) in studies on the Carpenter (2013) meta-analysis than for the studies made after.

Method

### Open-science, replicability, and our current study

We preregistered our analysis, following PRISMA (Moher et al., 2009) checklist and made available all our data and our analysis in R/Rmarkdown in an OSF (link = https://osf.io/8eqa5/). R packages used can be found in supplementary.

Literature search

We systematically searched Google Scholar (for suitability for meta-analyses see Gehanno et al., 2013; Martín-Martín et al., 2018; Walters, 2007) with the following term but you are free, as Carpenter did in 2013. We provide an overview of the search process in Figure 1. The database searches achieved 1760 hits. We also searched articles by scanning reference sections of

Figure 1

Meta-analysis flow diagram (adapted from PRISMA 2020)
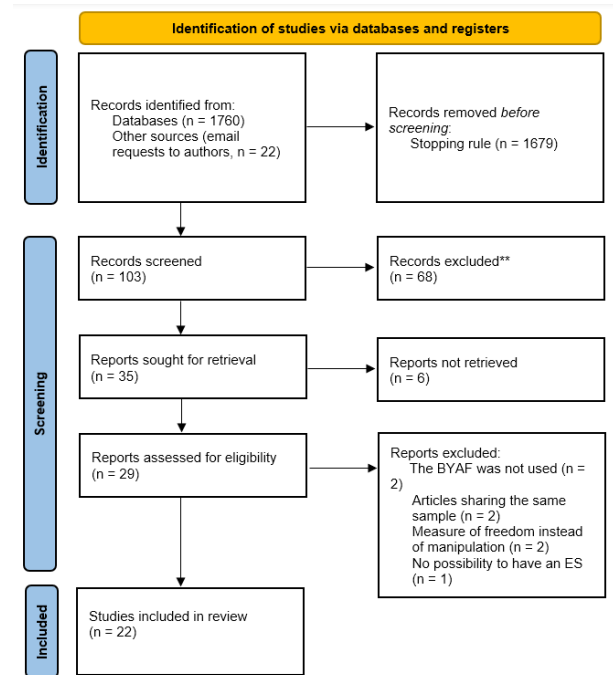
found articles and using the "related articles" and "cited by" options in Google Scholar. Based on reviewer feedback, we asked for unpublished studies in the ADRIPS, EADM, and EASP social networks, without any additional results.

After adjusting for duplicates, 81 sources remained. To minimize possible potential publication bias, we contacted all identified authors in person and requested unpublished manuscripts. We were provided with twenty-two additional articles leading to a total of 103 sources. All abstracts, tables, and results sections of empirical sources were scanned to assess their relevance. After this step, 29 articles remained as potentially includable articles. Our eligibility criterion is the use of the "But you are free" technique with a direct measure of compliance. We only include experimental designs, with a clear contrast between the BYAF technique and a control group, with an asking being saying "yes/no", money, clicking on a button online, or sending a postal mail. We exclude studies 1) that do not measure direct compliance or are using a scale to measure the strength of compliance, 2) without a control group, and that contrasts the BYAF technique with another technique and 3) that do not provide the exact term for the BYAF technique, for whom the term is disconnected/too far away

from the term "but you are free". Finally, we exclude studies with missing statistics or statistics that are not reported: Studies that do not report crucial measures such as the number of participants or standard needed for calculating the effect size deviation will be excluded from the sample. We briefly read through all articles to examine whether they met our inclusion criteria. A total of 7 articles were qualified for the exclusion, leading to a total sum of 22 identified articles with codable data. Finally, a total of 52 samples were included in this meta-analysis leading to a sum of 74 effect sizes. We provided a list of all included experiments in Table 1. We used a data extraction sheet that was already successfully used in other meta-analyses (e.g., Fillon et al., 2021; Yeung et al., 2021). The coding process for the pre-tests was completed by two coders to ensure high inter-rater reliability. We documented and reported all decisions in detail. After testing, one review author extracted all data and provided detailed information about coding decisions. A second author verified the coding. Disagreements were resolved by discussion between the two authors. All coding decisions were documented in the extraction sheet. We added in OSF available raw data and emails with authors. We documented in column "source" the extraction of data.

Coding

Included studies

We included a total of 52 experiments with a total of 19528 participants. The final sample consists of 18 published and 4 unpublished studies. Most studies were conducted in a face-to-face experimental design, in the street; others were made online, via an online video game or by email, phone, or postal letters. An overview of all included studies is provided in Table 1.

Analysis

We ran our analysis in R. We used the following meta-analysis related packages to conduct our analyses: metafor, psych, compute.es, MBESS, MAd, powerAnalysis, metaforest, esc, metaviz, puniform, zcurve (see supplementary for the whole R packages used). Given the range of different types of studies and designs, we expected heterogeneity in the sample to be relatively high. Therefore, a random-effects model was used. We coded the sheet with the total number of participants in each group (experimental via the BYAF technique, control) and the number of participants who comply in each group. In most cases, the numbers were provided but for some, we computed them from the test available. All conversions and coding decisions were documented. We preregistered to use cohen's d as effect size but used

Hedges' g instead because it corrects for low sample size (Delacre et al., 2021). We produced forest plots of the effect size distribution. A meta-analysis examined the overall main effect of the bias; a meta-regression was conducted to assess the impact of the described moderators. Statistical heterogeneity was determined using the Tau^2 test and quantified using I^2, which represents the percentage of the total variation in a set of studies that is due to heterogeneity (Higgins, 2003). This yielded a point estimate, confidence interval, and p-value, along with statistics for heterogeneity, assessed using the Q-statistics, and the I^2 statistic. We detected significant heterogeneity and therefore proceeded to explore potential moderators. We also performed analyses for the presence of publication bias, including funnel plots and statistical tests for publication bias (publication status as a moderator) and funnel plot asymmetry tests (Trim-and-fill method, rank correlation test, Egger's unweighted regression symmetry test, etc.). Finally, we tested for robustness via the Graphical Display of Study Heterogeneity (Gosh) and plotted a Z-curve to estimate replicability.

Moderator analyses

We tested subgroups and moderators using a comparison of fixed-effects meta-analysis models. Most of our hypotheses are exploratory; we tested the type of request and immediate or delayed as confirmatory, since they were already studied in the Carpenter (2013) meta-analysis. For the other moderators, we conducted exploratory analyses.

Results

The But-you-are-free main effect

In an analysis of all studies on the impact of the BYAF effect on compliance, we found an effect of $g = 0.44$ [0.36, 0.51]. We found considerable heterogeneity (Q (73) = 271.67, $p < .001$, $I^2 = .80.7\%$) in the observed effect sizes. The variation in effect-sizes was greater than would be expected from sampling error alone, indicating that moderator variables might be accountable for the variance in the effects. A meta-analysis forest plot is provided in Figure 2.

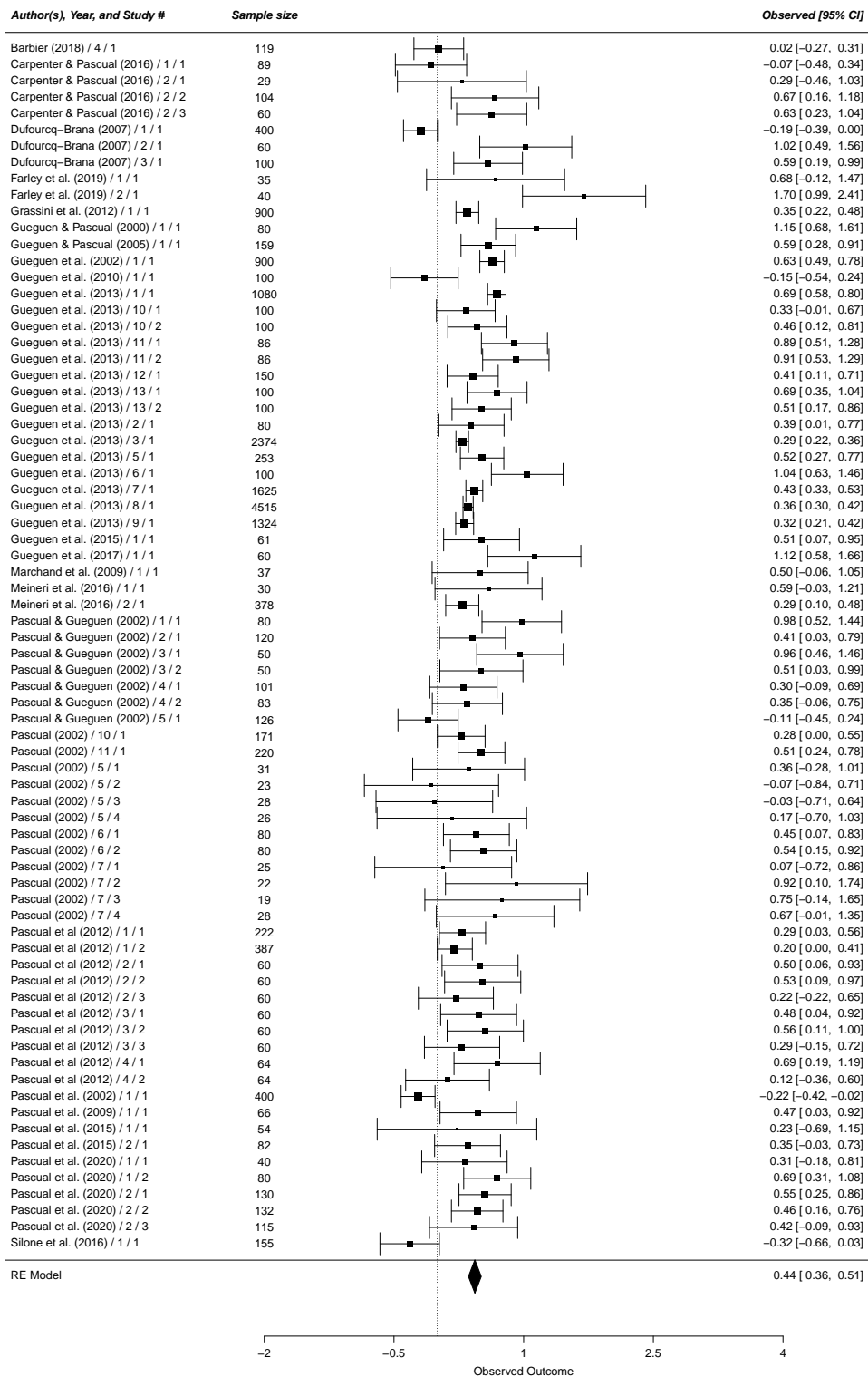### Study design and measures as moderators

We summarized all moderator findings in Table 2. Overall, the only exploratory moderator that has an impact on the BYAF effect was the type of interactivity, as face-to-face interactivity has a significantly higher number of compliers than the others combined (email, phone, postal letter, and internet). On the other side,

Table 1

All experiments included in the meta-analysis

	Article	N	Interactivity	Culture	Published
1	Barbier (2018)	422	Internet	France	No
2	Carpenter & Pascual (2016)	131	Face-to-face	USA	Yes
3	Carpenter & Pascual (2016)	320	Face-to-face	France	Yes
4	Carpenter & Pascual (2016)	240	Face-to-face	Norway	Yes
5	Dufourcq-Brana (2007)	400	Email	France	No
6	Dufourcq-Brana (2007)	60	Face-to-face	France	No
7	Dufourcq-Brana (2007)	100	Face-to-face	France	No
8	Farley et al. (2019)	45	Face-to-face	USA	Yes
9	Farley et al. (2019)	40	Face-to-face	USA	Yes
10	Grassini et al. (2012)	900	Email	France	Yes
11	Guéguen & Pascual (2000)	80	Face-to-face	France	Yes
12	Guéguen & Pascual (2005)	159	Face-to-face	France	Yes
13	Guéguen et al. (2002)	600	Email	France	Yes
14	Guéguen et al. (2010)	100	Face-to-face	France	Yes
15	Guéguen et al. (2013)	2160	Face-to-face	France	Yes
16	Guéguen et al. (2013)	160	Face-to-face	France	Yes
17	Guéguen et al. (2013)	4421	Face-to-face	France	Yes
18	Guéguen et al. (2013)	400	Face-to-face	France	Yes
19	Guéguen et al. (2013)	100	Face-to-face	France	Yes
20	Guéguen et al. (2013)	2608	Phone	France	Yes
21	Guéguen et al. (2013)	4515	Email	France	Yes
22	Guéguen et al. (2013)	2230	Postal letter	France	Yes
23	Guéguen et al. (2013)	400	Postal letter	France	Yes
24	Guéguen et al. (2013)	344	Face-to-face	France	Yes
25	Guéguen et al. (2013)	300	Face-to-face	France	Yes
26	Guéguen et al. (2013)	400	Face-to-face	France	Yes
27	Guéguen et al. (2015)	120	Face-to-face	France	Yes
28	Guéguen et al. (2017)	60	Face-to-face	France	Yes
29	Marchand et al. (2009)	74	Face-to-face	France	Yes
30	Meineri et al. (2016)	60	Face-to-face	France	Yes
31	Meineri et al. (2016)	649	Face-to-face	France	Yes
32	Pascual & Guéguen (2002)	80	Face-to-face	France	Yes
33	Pascual & Guéguen (2002)	120	Face-to-face	France	Yes
34	Pascual & Guéguen (2002)	200	Face-to-face	France	Yes
35	Pascual & Guéguen (2002)	306	Face-to-face	France	Yes
36	Pascual & Guéguen (2002)	126	Face-to-face	France	Yes
37	Pascual (2002)	181	Face-to-face	France	No
38	Pascual (2002)	320	Face-to-face	France	No
39	Pascual (2002)	167	Face-to-face	France	No
40	Pascual (2002)	306	Face-to-face	France	No
41	Pascual (2002)	220	Face-to-face	France	No
42	Pascual et al. (2012)	609	Face-to-face	France, Ivory Coast	Yes
43	Pascual et al. (2012)	360	Face-to-face	France, Romania, Russia	Yes
44	Pascual et al. (2012)	360	Face-to-face	France, Romania, Russia	Yes
45	Pascual et al. (2012)	128	Face-to-face	France, China	Yes
46	Pascual et al. (2002)	400	Email	France	Yes
47	Pascual et al. (2009)	120	Face-to-face	France	Yes
48	Pascual et al. (2015)	60	Face-to-face	France	Yes
49	Pascual et al. (2015)	160	Face-to-face	France	Yes
50	Pascual et al. (2020)	314	Face-to-face	France, China	Yes
51	Pascual et al. (2020)	788	Face-to-face	France, Moldavia, Tunisia	Yes
52	Silone et al. (2016)	155	Postal letter	France	Yes

Figure 2

Meta-analysis forest plot for all studies

Author(s), Year, and Study #	Sample size		Observed [95% CI]
Barbier (2018) / 4 / 1	119		0.02 [−0.27, 0.31]
Carpenter & Pascual (2016) / 1 / 1	89		−0.07 [−0.48, 0.34]
Carpenter & Pascual (2016) / 2 / 1	29		0.29 [−0.46, 1.03]
Carpenter & Pascual (2016) / 2 / 2	104		0.67 [0.16, 1.18]
Carpenter & Pascual (2016) / 2 / 3	60		0.63 [0.23, 1.04]
Dufourcq–Brana (2007) / 1 / 1	400		−0.19 [−0.39, 0.00]
Dufourcq–Brana (2007) / 2 / 1	60		1.02 [0.49, 1.56]
Dufourcq–Brana (2007) / 3 / 1	100		0.59 [0.19, 0.99]
Farley et al. (2019) / 1 / 1	35		0.68 [−0.12, 1.47]
Farley et al. (2019) / 2 / 1	40		1.70 [0.99, 2.41]
Grassini et al. (2012) / 1 / 1	900		0.35 [0.22, 0.48]
Gueguen & Pascual (2000) / 1 / 1	80		1.15 [0.68, 1.61]
Gueguen & Pascual (2005) / 1 / 1	159		0.59 [0.28, 0.91]
Gueguen et al. (2002) / 1 / 1	900		0.63 [0.49, 0.78]
Gueguen et al. (2010) / 1 / 1	100		−0.15 [−0.54, 0.24]
Gueguen et al. (2013) / 1 / 1	1080		0.69 [0.58, 0.80]
Gueguen et al. (2013) / 10 / 1	100		0.33 [−0.01, 0.67]
Gueguen et al. (2013) / 10 / 2	100		0.46 [0.12, 0.81]
Gueguen et al. (2013) / 11 / 1	86		0.89 [0.51, 1.28]
Gueguen et al. (2013) / 11 / 2	86		0.91 [0.53, 1.29]
Gueguen et al. (2013) / 12 / 1	150		0.41 [0.11, 0.71]
Gueguen et al. (2013) / 13 / 1	100		0.69 [0.35, 1.04]
Gueguen et al. (2013) / 13 / 2	100		0.51 [0.17, 0.86]
Gueguen et al. (2013) / 2 / 1	80		0.39 [0.01, 0.77]
Gueguen et al. (2013) / 3 / 1	2374		0.29 [0.22, 0.36]
Gueguen et al. (2013) / 5 / 1	253		0.52 [0.27, 0.77]
Gueguen et al. (2013) / 6 / 1	100		1.04 [0.63, 1.46]
Gueguen et al. (2013) / 7 / 1	1625		0.43 [0.33, 0.53]
Gueguen et al. (2013) / 8 / 1	4515		0.36 [0.30, 0.42]
Gueguen et al. (2013) / 9 / 1	1324		0.32 [0.21, 0.42]
Gueguen et al. (2015) / 1 / 1	61		0.51 [0.07, 0.95]
Gueguen et al. (2017) / 1 / 1	60		1.12 [0.58, 1.66]
Marchand et al. (2009) / 1 / 1	37		0.50 [−0.06, 1.05]
Meineri et al. (2016) / 1 / 1	30		0.59 [−0.03, 1.21]
Meineri et al. (2016) / 2 / 1	378		0.29 [0.10, 0.48]
Pascual & Gueguen (2002) / 1 / 1	80		0.98 [0.52, 1.44]
Pascual & Gueguen (2002) / 2 / 1	120		0.41 [0.03, 0.79]
Pascual & Gueguen (2002) / 3 / 1	50		0.96 [0.46, 1.46]
Pascual & Gueguen (2002) / 3 / 2	50		0.51 [0.03, 0.99]
Pascual & Gueguen (2002) / 4 / 1	101		0.30 [−0.09, 0.69]
Pascual & Gueguen (2002) / 4 / 2	83		0.35 [−0.06, 0.75]
Pascual & Gueguen (2002) / 5 / 1	126		−0.11 [−0.45, 0.24]
Pascual (2002) / 10 / 1	171		0.28 [0.00, 0.55]
Pascual (2002) / 11 / 1	220		0.51 [0.24, 0.78]
Pascual (2002) / 5 / 1	31		0.36 [−0.28, 1.01]
Pascual (2002) / 5 / 2	23		−0.07 [−0.84, 0.71]
Pascual (2002) / 5 / 3	28		−0.03 [−0.71, 0.64]
Pascual (2002) / 5 / 4	26		0.17 [−0.70, 1.03]
Pascual (2002) / 6 / 1	80		0.45 [0.07, 0.83]
Pascual (2002) / 6 / 2	80		0.54 [0.15, 0.92]
Pascual (2002) / 7 / 1	25		0.07 [−0.72, 0.86]
Pascual (2002) / 7 / 2	22		0.92 [0.10, 1.74]
Pascual (2002) / 7 / 3	19		0.75 [−0.14, 1.65]
Pascual (2002) / 7 / 4	28		0.67 [−0.01, 1.35]
Pascual et al (2012) / 1 / 1	222		0.29 [0.03, 0.56]
Pascual et al (2012) / 1 / 2	387		0.20 [0.00, 0.41]
Pascual et al (2012) / 2 / 1	60		0.50 [0.06, 0.93]
Pascual et al (2012) / 2 / 2	60		0.53 [0.09, 0.97]
Pascual et al (2012) / 2 / 3	60		0.22 [−0.22, 0.65]
Pascual et al (2012) / 3 / 1	60		0.48 [0.04, 0.92]
Pascual et al (2012) / 3 / 2	60		0.56 [0.11, 1.00]
Pascual et al (2012) / 3 / 3	60		0.29 [−0.15, 0.72]
Pascual et al (2012) / 4 / 1	64		0.69 [0.19, 1.19]
Pascual et al (2012) / 4 / 2	64		0.12 [−0.36, 0.60]
Pascual et al. (2002) / 1 / 1	400		−0.22 [−0.42, −0.02]
Pascual et al. (2009) / 1 / 1	66		0.47 [0.03, 0.92]
Pascual et al. (2015) / 1 / 1	54		0.23 [−0.69, 1.15]
Pascual et al. (2015) / 2 / 1	82		0.35 [−0.03, 0.73]
Pascual et al. (2020) / 1 / 1	40		0.31 [−0.18, 0.81]
Pascual et al. (2020) / 1 / 2	80		0.69 [0.31, 1.08]
Pascual et al. (2020) / 2 / 1	130		0.55 [0.25, 0.86]
Pascual et al. (2020) / 2 / 2	132		0.46 [0.16, 0.76]
Pascual et al. (2020) / 2 / 3	115		0.42 [−0.09, 0.93]
Silone et al. (2016) / 1 / 1	155		−0.32 [−0.66, 0.03]
RE Model			0.44 [0.36, 0.51]

the two confirmatory moderators had a significant effect, as we found that a face-to-face interaction led to a stronger effect than the other forms of interactivity, and a direct request led to a stronger effect than a delayed request.

Subject gender

We hypothesized that the BYAF technique would increase compliance to a higher degree with women than with men. While we found a slightly larger effect size of the BYAF technique for women, this difference was not statistically significant.

Confederate gender

We hypothesized that the BYAF technique would increase compliance to a higher degree with women than with men confederates. We did not find support for this hypothesis, as the test for the difference was non-significant. We also performed an ANOVA on the confederate and subject gender moderators to find if there might be an interaction effect. The ANOVA revealed no statistically significant interaction effect (Q (3) = 2.18, $p = 0.54$).

Culture

We hypothesized that the BYAF technique would increase compliance to a higher degree in individualistic cultures than in collectivistic cultures. Our results indicate a higher effect size of the BYAF technique in individualistic culture than collectivistic, but the result is not significant.

Interactivity

We hypothesized that the BYAF technique would increase compliance to a higher degree in Face-to-face interaction than the other types of interaction. Our results indicate a higher and significant effect size of the BYAF technique with the face-to-face interaction than the other, yet we caution against drawing any general conclusions from these findings as we did not find enough effect sizes for the "other" moderators. For example, we only collected one effect size for the use of the technique by phone.

Freedom evocation

We hypothesized a stronger effect of the BYAF technique with the exact term "but you are free" than other terms. On the contrary, our results indicate a higher effect of the combined other framing, while the effect is not significant.

Carpenter's analysis

We hypothesized a stronger effect size via the coding of the Carpenter's (2013) meta-analysis than the effect sizes found in the experiments made after the Carpenter analysis. We did not find any differences between the studies made before and after Carpenter's analysis, as the average effect sizes are very similar.

Type of request

We hypothesized a higher number of compliers with the BYAF technique in a prosocial request than a selfish one. Our result tends to indicate the contrary, participants complied more with a selfish request than a prosocial request with the BYAF technique, but the effect is not significant.

Temporality

We hypothesized that the effect of the BYAF technique would be stronger for immediate requests and weaker for delayed ones. Our results corroborate the hypothesis; we found a stronger and significant effect for immediate requests ($g = 0.47$, 95% CI [0.41; 0.54]) than for delayed requests ($g = 0.25$, 95% CI [0.03, 0.47]).

Publication bias

We tested for the presence of publication bias using several methods, and a summary of publication bias analyses is provided in Table 3. We ran publication bias analyses on collapsed effect sizes by study, with one effect size per study. Point estimates are consistent, and methods that produce confidence intervals show substantial overlap in confidence intervals for each method. The range of estimates goes from 0.25 to 0.56. The trim and fill method indicates an asymmetry of the funnel with 17 studies missing on the left side, confirmed with a significant Egger's regression test. The asymmetry of a funnel plot can be caused by two effects: publication bias or other factors (e.g., poor methodological quality, true heterogeneity, artefactual, or chance; Egger et al., 1997). The distinction between publication bias and other factors relies on where the missing studies are in the funnel plot. If the missing studies are in the significant area (i.e., the white area inside the funnel plot), it means that the meta-analysis lacks significant effect sizes, which are mainly due to other factors. If the missing studies are in the non-significant area (i.e., the darker areas of the funnel plot), it probably indicates a sign of publication bias. Based on the Funnel plot (Figure 3) and the Trim-and-Fill plot (Figure 4), our results indicate the presence for both signs,

Table 2

Moderator analysis of the but you are free technique

Moderator	k	N	Mean g	95% CI	*Difference*	p
Subject gender						
Woman	38	9316	0.48	[0.40, 0.56]		
Man	41	8008	0.42	[0.35, 0.50]	-0.059 [-0.17, 0.05]	.28
Confederate gender						
Woman	50	4355	0.45	[0.36, 0.54]		
Man	26	2048	0.41	[0.27, 0.55]	-0.04 [-0.21, 0.13]	.62
Culture						
Individualistic	65	18550	0.45	[0.37, 0.53]		
Collectivistic	9	978	0.32	[0.20, 0.44]	0.13 [-0.01, 0.28]	.08
Interactivity						
Face-to-face	64	9101	0.49	[0.42, 0.56]		
By e-mail	5	7115	0.19	[-0.13, 0.52]		
By phone	1	1625	0.43	[0.33, 0.53]		
By postal letter	2	1479	0.02	[-0.60, 0.64]		
By internet	2	208	-0.01	[-0.25, 0.23]		
Overall other than Face-to-face	10	10427	0.15	[-0.05, 0.36]	-0.34** [-0.55, -0.13]	.002
Freedom evocation						
« But you are free »	59	14069	0.42	[0.34, 0.50]		
Other	13	5218	0.53	[0.36, 0.71]	0.11 [-0.08, 0.30]	.26
Carpenter						
Before	54	16835	0.44	[0.35, 0.52]		
After	20	2693	0.43	[0.27, 0.60]	0.007 [-0.18, 0.19]	.95
Type of request						
Selfish	40	12603	0.50	[0.41, 0.60]		
Prosocial	34	6925	0.36	[0.26, 0.47]	0.14 [-0.005, 0.29]	.06
Temporality						
Immediate	63	10451	0.47	[0.41, 0.54]		
Delayed	12	9212	0.25	[0.03, 0.47]	0.23 [-0.004, 0.45]	.05

Note. k = number of samples; N = total number of individuals in k; mean g = average Hedge's g effect size, CI = lower and upper limits of 95% confidence interval, * $p < .05$, two-tailed, ** $p < .01$, two-tailed, *** $p < .001$, two-tailed.

as we found support for a lack of significant and non-significant studies. These results are strengthened by the Three-parameter selection model (3PSM) estimate, for which the likelihood ratio test is close to the significance threshold, which could indicate selective reporting (Hedges, 1992). In the case of inconsistencies between estimators, the 3PSM is a better indication (Carter et al., 2019), and, in our case, does not exclude a possible publication bias. Overall, while some estimators indicate a possible publication bias, the more robust test for high heterogeneity do not favor the possibility for selective reporting. But this result is accompanied by a possible problem of poor methodological quality leading to a (rather small) inflated effect, from a found effect of 0.44 to an estimated mean effect between 0.34 and 0.38. We ran a p-curve and p-uniform analysis

which respectively found an estimated $g = 0.41$ and $g = 0.38$. The p-uniform analysis found 45 significant effect sizes, and the p-curve analysis indicated presence of evidential values and no absence of evidential values (see supplementary for the P-curve Table). As requested by the editor, we ran a statcheck (Nuijten, 2018) on the statistics we used to retrieve the number of participants in each condition and found only one inconsistent result which did not affect the overall result.

Robustness

We did not pre-register an estimation of Robustness. Still, we ran a script to create a Graphical Display of Study Heterogeneity (GOSH) to assess the robustness of effect size found. We provide the R script in supplementary rather than in the Rmarkdown because of the
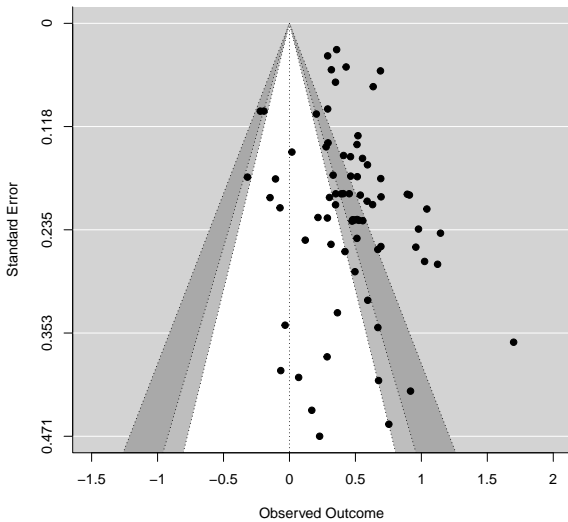
Table 3

Publication biases analyses results

Publication bias analysis method	Results and adjusted models
Three-parameter selection model	Likelihood Ratio Test: 3.39, p = .07 Adjusted Model: g = 0.38, 95% CI [0.26, 0.50]
PET	b = 0.34 [0.25, 042], p <.001
PEESE	b = 0.36 [0.30, 0.42], p <.001
Puniform	Adjusted Model: g = 0.45, 95% CI [0.37 0.56], 45 significant
Henmi & Copas (2010)	Adjusted Model: g = 0.36, 95% CI [0.26, 0.51]
Trim and fill funnel plot asymmetry	17 studies missing on the left side.
Rank correlation test (Begg & Mazumdar, 1994)	Kendall's tau = 0.14, p = .09
Egger's regression test	z = 2.06, p = .04

Note. Values in parentheses indicate 95% confidence intervals [lower bound, upper bound].

Figure 3

Funnel plot for all studies

Figure 4

Trim-and-Fill funnel plot
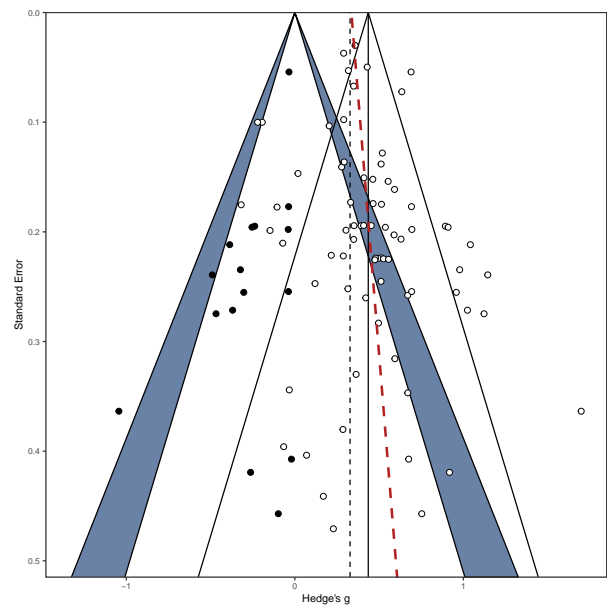
Note. The 17 missing studies are shown in black. We used the Trim-and-Fill method to see studies on the left with a random model, with the addition of the Egger regression test shown as the red line.

time consumption used in the analysis. On our recent computer, the analysis took between 3 and 4 hours. One test of robustness includes the leave-one-out analysis, a method of analysis (Olkin et al., 2012) made to see the influence of one effect size on heterogeneity. Another possibility is to estimate the influence of a subgroup in meta-analyses, which leads to a very high number of meta-analyses to perform to find the whole combination of effect sizes that could influence the robustness of the analysis. In fact, with 74 effect sizes found, it leads to 1.88×10^{22} meta-analysis, which makes the comparison

impossible. The GOSH makes the analysis graphical, by plotting one meta-analysis as a dot. If dots are homogeneously displayed, the effect found is robust, while if two or more clusters are found, it means that at least one subgroup influences too much the overall effect size found. Our GOSH plot can be found in Figure 5. The figure presented is in a homogenous circle form, show-

Figure 5

GOSH plot For Robustness
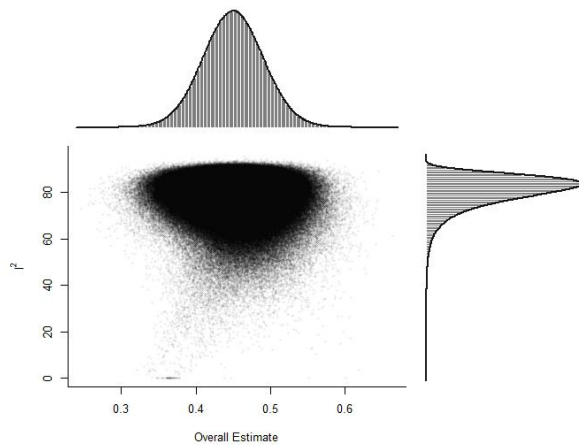
Note. The plot helps to see how heterogeneity varies between overall estimates for every left-out meta-analysis. We can see that for every meta-analysis, the overall estimate varies between 0.3 and 0.6, with heterogeneity between I^2=60% and $I^2 = 90\%$.

ing that all meta-analyses have an average estimate between 0.3 and 0.6, and heterogeneity between I^2=60% and $I^2 = 90\%$. We conclude that the meta-analysis estimate is robust to leave-out studies.

Z-curve analysis

Based on feedback from a reviewer, we created a z-curve analysis (Figure 6, Bartoš and Schimmack, 2021). The Z-curve is a method for estimating publication bias and possibility of false positives. The observed discovery rate is of 45% (64 significant tests out of 141). The expected discovery rate, or the mean power before selection for significance, is of 6%. The expected replication rate, or the mean power after selection for significance, is of 73%. Thus, we see that the power of studies after selection for significance is far higher than before. This is a clear indication of publication bias with a high false positive risk.

Risk Of Bias 2 (ROB2)

As asked by the editor, we conducted a ROB2 check (J. Higgins, Thomas, et al., 2022; McGuinness and Higgins, 2021). We detailed the check by domain alongside the assessment in the spreadsheet. Overall, we found that nearly 40% of studies did not randomize or declare the randomization of the participants, nearly 40% lacked an explanation of missing data, and 50% had

Figure 6

Z-curve Analysis of the But-you-are-free effect (expectation maximization, EM method)
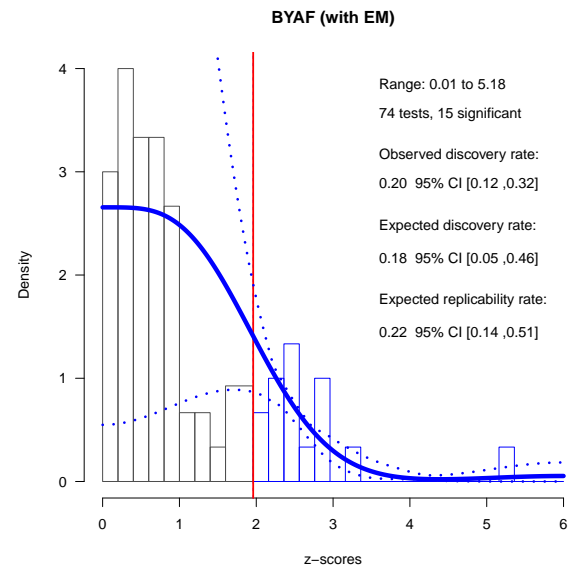
a bias in the measurement of the outcome because we cannot trust the studies made with Guéguen's students (see Brown, 2020). We found no bias due to deviation from the intended intervention because all interventions were straightforward, with the measurement of the direct behavior. Finally, no study declared a pre-planification (see Figure 7).

After conducting an overall risk of bias, we created a Traffic-light plot visualizing the risks by study (see Figure 8). The complete Plot is in supplementary materials on OSF. We found only seven studies on low-risk and decided to run another meta-analysis on these studies.

Our seven studies indicated no effect of the BYAF technique, with a g = 0.11, 95% CI [-0.18; 0.40]. The heterogeneity was huge, with a $I^2 = 95\%$. A forest plot of the effect can be found in Figure 9. Based on an exchange with the editor, we conducted a third meta-analysis, including all studies with an overall rate of "low risk" and "some concerns". The result of the meta-analysis is g = 0.38, 95% CI [0.27; 0.49] and $I^2 = 84\%$.

Discussion

We conducted three meta-analyses for the BYAF technique. We tested several moderators and found support for a contextual effect of the technique on compliance. Including all studies, we found a direct medium effect of the BYAF technique (g = 0.44) consistent across most of

Figure 7

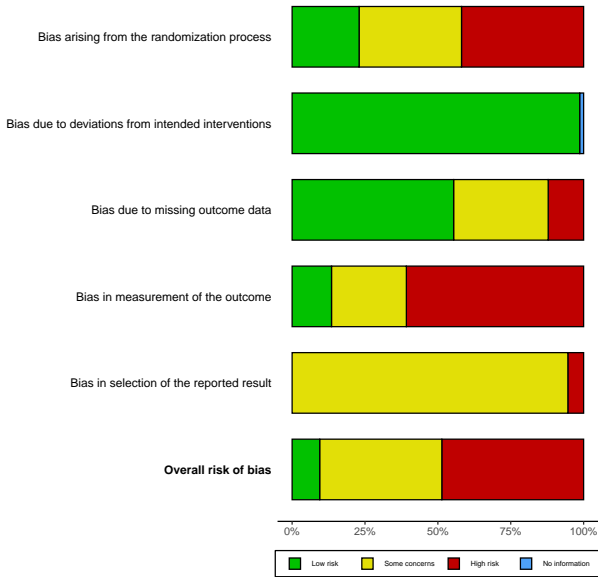*Risk of bias in studies included in our meta-analysis*

Figure 8

Traffic-light plot of the ten firsts studies included in our meta-analysis

our moderators. Excluding high-risk studies, the effect found was weaker ($g = 0.38$) and nonexistent with only low risk of bias studies ($g = 0.11$, CI including the null).

Confirmatory moderators

Type of request

We initially hypothesized that the efficacy of the BYAF technique might be higher for prosocial requests than selfish requests, as Carpenter (2013) first hypothesized. In his meta-analysis, he did not find evidence that prosocial requests were associated with a higher level of compliance. With the addition of new effect sizes based on new experiments, we also did not find a significant dif-

ference, but our difference is now in the other direction: our results indicate a non-significant higher effect size for selfish requests. This result, while surprising, might be confounded with other moderators. Indeed, selfish requests are often made face-to-face and immediately, two conditions with high averaged effect sizes, while prosocial requests were often made indirectly and in delayed condition, two conditions with lower averaged effect sizes. In the prosocial condition, the effect size found ($g = 0.36$) was medium, indicating that this moderator does not play a fundamental role in the effectiveness of the BYAF technique: it might work independently of this contextual effect.

Temporality

Our moderator analysis revealed that the effect of the BYAF technique is stronger for immediate rather than for delayed requests, while being at the threshold for significance ($\alpha = .05$, $p = .05$). This finding is in accordance with our hypothesis, and the findings of Carpenter (2013). Indeed, the effect found in the immediate condition was medium to strong ($g = 0.47$) and weak in the delayed condition ($g = 0.25$). This finding is not surprising, since we only added two effect sizes to the delayed condition regarding Carpenter's meta-analysis. Most recent works in the BYAF literature are made via the immediate condition. The effectiveness of the BYAF technique is impacted by the temporality moderator: once the demand is delayed, we cannot be sure that the BYAF technique can be effective, which shows the importance for the participant to be directly linked to the confederate.

Exploratory moderator
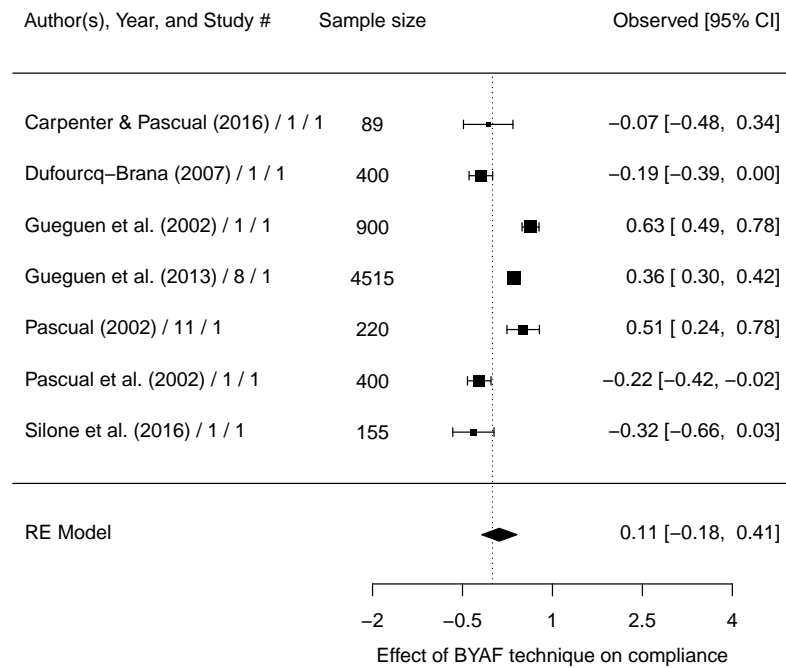
Subject and confederate gender

We hypothesized that the BYAF technique can make women comply more than men and that individuals would comply more with a woman confederate. We found no support for a gender effect of moderation. Indeed, across the four conditions, the effect sizes remain constant (between $g = 0.41$ and $g = 0.48$). Also, the result from the ANOVA reveals no interaction effect: the gender of the individual does not interact with the gender of the confederate.

Culture

We hypothesized that the BYAF technique could be stronger in individualistic countries than in collectivistic countries. Our results could possibly corroborate this hypothesis. However, the p-value is not significant,

Figure 9

Forest plot of "low risk" studies included in our meta-analysis

Author(s), Year, and Study #	Sample size	Observed [95% CI]
Carpenter & Pascual (2016) / 1 / 1	89	−0.07 [−0.48, 0.34]
Dufourcq–Brana (2007) / 1 / 1	400	−0.19 [−0.39, 0.00]
Gueguen et al. (2002) / 1 / 1	900	0.63 [0.49, 0.78]
Gueguen et al. (2013) / 8 / 1	4515	0.36 [0.30, 0.42]
Pascual (2002) / 11 / 1	220	0.51 [0.24, 0.78]
Pascual et al. (2002) / 1 / 1	400	−0.22 [−0.42, −0.02]
Silone et al. (2016) / 1 / 1	155	−0.32 [−0.66, 0.03]
RE Model		0.11 [−0.18, 0.41]

Effect of BYAF technique on compliance

and we only found 9 effect sizes for participants in collectivistic countries, which limits our possibility of explanation. Nonetheless, we found that the BYAF technique can be more effective in an individualistic setting. This might be in part due to the easiness for people in individualistic countries to be reactant to the asking, and more effectiveness of the BYAF technique to lower the reactance in this situation. Other mental processes might be active in individualistic countries to influence people not to comply, but they remain unknown.

Interactivity

We hypothesized that face-to-face interaction would lead to more compliance with the BYAF effect than other types of interaction. Overall, we found a significant difference in this direction: the face-to-face interaction found the highest average effect size ($g = 0.51$). In more detail, we found that phoning could be a good way to exercise the BYAF technique with a medium effect size ($g = 0.43$), but we only found one study with this type of interaction. E-mail can also be an effective way to use the BYAF technique, but the effect size

found was considerably lower ($g = 0.19$) and included the null. We call for further examination of this condition of interactivity, since the results are not clear. For the other types of interactivities (i.e., postal letter, internet) we found no effect of the BYAF technique, but we are limited by the number of studies included, with only 2 effect sizes found for each condition. Overall, we found a significant difference between the face-to-face interaction and the others, but we cannot draw a definitive conclusion due to too few effect sizes in the other conditions.

Freedom evocation

The goal of this moderator was to understand if the exact term "But you are free" was necessary for the effect to appear. We found that it was not the case, as the effects found were not different between the exact term and others. The combination of the other terms leads to a higher non-significant average effect size, signaling a possible more effective way or term to induce compliance than the standard term "but you are free". But what were the other types of evocation used that give

the highest effect sizes? Given the forest plot (See supplementary metarials), we see that at least three studies give a very high effect size. In the first (Farley et al., 2019, study 2, $g = 1.70$) the confederate added the term "feel free to say no". In the second (Guéguen et al., 2013, study 11, $g = 0.91$), the confederate added the term "Do as you wish" and in the third (Pascual and Guéguen, 2002, study 7, $g = 0.75$), the confederate added the term "you are not obliged". In comparing the three terms, we do not find any patterns leading to a meaningful conclusion about how they lead to a stronger effect of the BYAF technique. The only common point between the three studies is that they have very few participants (respectively 40, 86, and 19) leading to a probable overestimate of the effect size.

Before and After Carpenter (2013)

Finally, we wanted to see if any differences were made between the studies before and after the Carpenter (2013) analysis to see if they lead to a different effect size found. Carpenter (Carpenter, 2013) found an average effect size of $r = .13$. Once our overall effect size ($g = 0.45$) was transformed in correlation, we found an $r = 0.22$ of the technique, two times higher than Carpenter found. This result still holds for the analysis we made of the identical dataset used by Carpenter. Why do we have so much difference? We found several errors in the Carpenter (Carpenter, 2013) analysis. For example, Carpenter used one experiment (Dufourcq-Brana, 2007) two times. Also, Carpenter made ambiguous and not reported decisions in his study. For example, for the experiments with two measures (e.g., Guéguen et al., 2002; Marchand et al., 2009; Pascual, 2002), he decided to take one of them and did not make transparent the reason why. In our analysis, we decided to merge them except if one variable is not includable as reported in our preregistration. We found several errors in the original papers, some tests were not compatible with the reported number of participants (e.g., Guéguen et al., 2013; Pascual et al., 2002, study 10). All the discrepancies found are made open in the commentary columns in the dataset. For the publication bias section, Carpenter only used the Trim-and-Fill technique, leading to no missing studies. We do not know how researchers used the algorithm (bilateral or left-centered, as recommended) and the Trim-and-Fill plot is not available. Also, researchers did not report the heterogeneity (I^2 or tau^2) found in the article, while giving the percentage of variability explained by sampling error. They still found that the BYAF technique only accounts for 22% of the variation, a condition in which the trim-and-fill tool alone might not be sensitive (Carter et al., 2019). Thus, we think that the use of this only publication bias estimator is not enough to assess the credibility of the effect size found. Finally, with the use of the Trim-and-Fill, Carpenter (2013) found an overall corrected effect size of $r = 0.04$. We found that: 1) the averaged effect size found was much higher than the one reported by Carpenter (2013) with the overall sample, 2) the averaged effect size did not differ from before and after the analysis made in 2013, 3) there was a lack of transparency of the choices made in 2013, leading to some errors and curious effect sizes taken into account and 4) no enough assessment of possible publication bias leading to think that the effect size found was more meaningful than it possibly is.

Implication

With all studies included, we found a medium effect size, but only one meaningful moderator, as the BYAF technique works better in the face-to-face condition than in others, with possible covariates. Also, we have several publication bias estimators flagging possible problems in relation to the experiments on this technique. We did not find that temporality is important to the effectiveness of the BYAF technique. More surprisingly, we did not find that subject and confederate genders were important. Also, we did not find differences between a selfish and a prosocial request and found quite the contrary, as selfish requests were more prone to the BYAF technique than prosocial. Our results indicate that participants seem not to process the request more carefully for a selfish request than for a prosocial one. The interactivity moderator was significant, but with too few studies for most modalities, and the merging of them can mislead our results. Finally, culture was barely significant, with far more participants from France than the other countries, and we cannot be sure that the effect is clearly related to culture and not to country and/or confederates in these countries. Overall, we did not find any consistent evidence for possible moderators. We have few publication bias estimators that indicate a possibility of publication bias. We found a little asymmetry in our funnel plot, for significant and non-significant results, via different techniques. According to Egger et al. (1997), four possibilities are to consider for this asymmetry: selection bias, poor methodological quality, true heterogeneity, and artefactual. For the selection bias, it might be possible to have location and language bias. For example, most of the original experiments on compliance were said to be made in the same street in the same city (Vannes, France), and others in Bordeaux (France). Also, it is possible that the "but you are free" and "vous-êtes-libre-de" do not have the same meaning, most importantly once translated into the language in collec-

tivistic countries. We found selective reporting in the Carpenter (2013) meta-analysis (reported in a commentary in the dataset). By looking at the original articles, we found lacking and inconsistent data we had to ask the author (we reported their answer in OSF). We also found several poor methodological qualities: a lot of the studies we found have very low power when compared to the effect size found in Carpenter's analysis. For example, with a correlation of $r = 0.13$ transformed to a $d = 0.26$, a power of 95%, equal number of participants in each group, α of 5%, and a one-tailed test (since we do not want the control group being more effective than the BYAF), we would need 321 participants per group to have a chance to detect the effectiveness of the technique (See supplementary materials for more details). In the forest plot of articles published after the Carpenter (2013) analysis (see supplementary materials), we find that only one article (i.e., Grassini et al., 2012) has the necessary power to detect an effect. Unfortunately, this experiment was made via e-mail and does not give us information about the standard face-to-face use of the technique, eliminating possible unknown covariates linked to the use of an online store. Overall, given the smallest effect size of interest of $r = .13$, no studies conducted are enough powered to ensure that the effect of the BYAF technique leads to compliance. For true heterogeneity, we see that the confidence interval is mostly high, due to too low sample size.
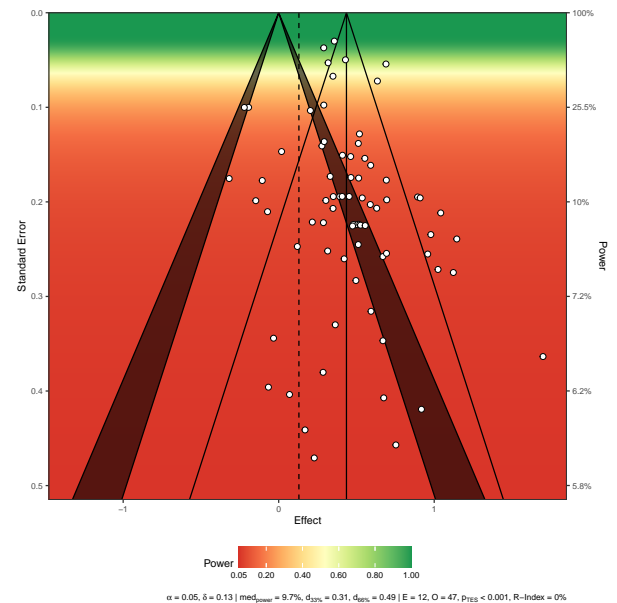
Limitations

Sample size and Power

In the first published paper on the BYAF technique, researchers employed 20 participants per condition (Guéguen and Pascual, 2000). Afterward, Carpenter found a very low effect size for the BYAF technique, which implies the need for a large sample ($n = 321$ per condition for 95% power, $n = 240$ per condition for 80% power and an alpha of 0.05, as shown in the implication section). In the last experiment on the subject (Farley et al., 2019), researchers assigned 25 participants to the BYAF group, and 20 to the control group. In between, we found no studies with enough participants in sample size to possibly detect an effect if the effect exists. Low sample size is a major concern for the possibility to put into evidence the effectiveness of the BYAF technique. To see the power of each study in the meta-analysis, we performed a power test (Figure 10). We set the test with a $r = 0.13$ and alpha = 0.05. The redder the area, the less power, the greener, the more. We found 5 studies in the green area and only two in the yellow area. The average power is 9.70% and the replicability index 0% which means that we have less than 10% of chance to

Figure 10

Power test of the articles published after Carpenter's (2013) analysis

$\alpha = 0.05, \delta = 0.13 \mid \text{med}_{power} = 9.7\%, d_{33\%} = 0.31, d_{66\%} = 0.49 \mid E = 12, O = 47, p_{TES} < 0.001, R\text{–Index} = 0\%$

Note.. We set alpha to 0.05 and an effect size to r = 0.13, the effect size found by Carpenter (2013). The redder the area, the less power, the greener, the more. We found no studies in the green area and only one in the yellow area. The average power is 9.20% and the replicability index 0% which means that we have less than 10% of chance to reject H0 when there is a true effect, and no chance at all to replicate one study (see Motyl et al., 2017 for R-index).

reject H0 when there is a true effect, and no chance at all to replicate one study (see Motyl et al., 2017 for R-index). Also, the Z-curve showed a very low discovery rate of 6%.

Guéguen's work

One main reason for conducting this meta-analysis was to see how reliable the effect of the But You Are Free technique was. One major limitation of the present meta-analysis is that nearly all the studies using this technique had Guéguen's authorship or were made by a Ph.D. student or close collaborator of Guéguen. We tried to make a meta-analysis without Guéguen's name and found a similar effect size of $g = 0.48$ [0.28; 0.68] with a total $N = 1010$ and $n = 457$ participants from Pascual and collaborators (2021) study. Most implausible Odds ratio comes from Guéguen's study, as we found Odds higher than 5 and some close to 10, with the huge exception of Farley and collaborators (2019)

whose results were higher than $OR = 23$, mostly because of a lack of power (the data were 19/20 compliers in the BYAF condition, 9/20 in the control condition). We checked these studies using the ROB2 tool and found that they were problematic for many reasons such as no randomization, most confederates are young students aware of the experimentation, no pre-registration or curious way of selecting the participants. As Brown (2020) shows, we cannot trust the young student's confederates of Guéguen, because some fabricated their data. Finally, we conducted a "low risk" meta-analysis which showed no result of the effect. This result clearly questions the existence of the BYAF effect.

Limitations in moderators

We tried to test several moderators to reduce heterogeneity: when and how can we ensure that the BYAF is effective? For most of them, the numbers of experiments were particularly low. Also, when aggregating them, we did not find any differences between them, and even if we did, we could not draw a strong conclusion because these moderators are, for some, very different from each other. For the moderators with more than 10 effect sizes, we did not find any differences, and we cannot explain why the heterogeneity persists in the effects found. The only significant moderator we found was the one from our confirmatory hypothesis temporality, as we found that immediate requests are more effective than delayed ones. Finally, we did not find any evidence that moderators can diminish the heterogeneity of the BYAF technique, leading to the conclusion that: 1) we did not take into account the most important moderators, mostly because researchers failed to raise attention to them, 2) they are no important moderators in the BYAF technique, which contradicts the moderators found for others techniques (see Carpenter, 2013 for a review of some) or 3) the publication bias and/or the possibility of a truly random effect leads to an inflated effect size.

Culture

While it might be less important than the issues raised above, we cannot be sure that our simple dichotomy in individualistic versus collectivistic countries is well appropriate for this technique. Indeed, the BYAF technique might rely on subtle or important differences between countries, as some studies on cross-cultural psychology pointed out. For example, Boskł (2020) found that male complied far less to male in Poland, but not in England, based on a sociocultural model. We do not know if the distinction we made was the best possible and we cannot compare countries because, France

aside, all the others have experiments from only one study.

Approximation of effect

While we made transparent how we coded our effect, they are not all closely similar. Indeed, we have, for some studies, merged two conditions altogether to have a control group. In other studies, we took only one possible effect size, the one most closely related to the BYAF effect. Nonetheless, it might still be possible that our decisions lead to bias. This drawback may apply to almost every meta-analysis in empirical science, but we tried to improve transparency and complete reports to ensure having less bias possible. Finally, the ROB2 check was made only by the first author, and as transparent as it is, the coding is subjective and can lead to a selection of "low risk" studies different from another coder.

Direction for future research

The BYAF technique

Since the first meta-analysis, we did not find any analysis powered enough to detect an effect of the BYAF technique. The most conservative meta-analysis we made, with only the "low risk" studies, questioned the existence of the effect. The main direction to take is to make a well-powered study with the main and original context of the appearance of the technique, in face-to-face interaction with a request for a spare for the bus. This replication should be made by several confederates from the two genders, in many places across the world. Also, this amount of work can be done with collaborative replication to see how the effect varies across different contexts and environments. At best, the study should be a pre-registered experiment or registered report based on the minimum effect size of interest of $r = .13$, with a power of 90% and alpha of 0.05, leading to a required sample size of 616 participants. The ROB2 check helped us detail what could be needed for the quality of the study. First, one should carefully explain the randomization and selection of participants in the street. Confederates should not be aware of the experimental conditions and should not be the one who select the participants. Researchers also need to report the targeted subject who didn't decide to reply at all. Confederates should be of all ages, because only (very) young students were confederates in studies included in the present meta-analysis (in most studies, the mean age of confederates is close to 20 years old).

Moderators

Once a well-designed, highly powered study is made, it would be possible to investigate some moderators. For example, one type of moderator might be highly relevant. The interactivity (Face-to-face or in a more indirect setting), and the immediate or delayed moderators were significant, which means that the presence of the confederate might be a necessity for the BYAF technique to work. One direction is to do studies in an internet setting, leading to a refining of the BYAF technique to a nudge, an easy and cheap intervention in the choice architecture (Thaler and Sunstein, 2009), aiming to improve the acceptance to a request. The difference found between the internet and face-to-face setting could lead to a huge improvement to understand how the BYAF technique works. Also, the only study we have in an e-mail setting shows an effect size in the range of effects found in nudge theory (DellaVigna and Linos, 2020). Another direction to investigate is the respective impact of gender, age, and culture. We tried to investigate the impact of gender and found no effect, but we could not control for age, which can impact the relationship between gender of the confederate and participant, in the face-to-face setting. Having the help of confederates from multiple age and gender can help understand the impact of these social cues on the helping of others to the request. Also, once controlling for gender and sex, we can move on and enhance the theory by construing upon cultural variation, in different countries and cultures. Finally, we did not find any differences between the exact term of evocation "But-you-are-free" and others, now we propose not to pursue in this direction, until a well-powered preregistered replication of the initial effect is made.

Author Contact

Adrien Fillon, ERA Chair in Science and Innovation Policy & Studies (SInnoPSis),University of Cyprus, adrienfillon@hotmail.fr

Conflict of Interest and Funding

The authors declared no potential conflicts of interest with respect to the authorship and/or publication of this article. One author, who provided data and verified the coding, was an important author in the literature. However, the main coder was independent of the literature. Coding and verifications were made transparent in the Excel file and the source for the code is provided in the Excel file (column N).

The authors received no financial support for the research and/or authorship of this article.

Author Contributions

Adrien worked under the supervision of Lionel and Fabien at Aix-Marseille University for conducting the pre-registered meta-analysis. Adrien wrote the pre-registration, with verification and registration by Alexandre, Lionel, and Fabien. Adrien and Alexandre conducted the search of the literature, developed the coding scheme, and coded the articles. Adrien provided the RMarkdown code and analyses. Adrien summarized the methods and results and wrote the manuscript. Adrien, Alexandre, Lionel, and Fabien finalized the manuscript for submission.

Open Science Practices

This article earned the Preregistration+, Open Data and the Open Materials badge for preregistering the hypothesis and analysis before data collection, and for making the data and materials openly available. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process, including the open reviews, is published in the online supplement.

References

Bartoš, F., & Schimmack, U. (2021). Zcurve: An Implementation of Z-Curves. Retrieved November 26, 2021, from https://CRAN.R-project.org/package=zcurve

Boskł, P. (2020). Investigating the sociocultural models with cultural experiments: A Polish–English study on Request → Compliance in gender relations. *Asian Journal of Social Psychology*, *23*(2). https://doi.org/10.1111/ajsp.12408

Brehm, J. W. (1966). *A theory of psychological reactance*. Academic Press.

Brown, N. (2020). Nick Brown's blog: The Guéguen saga update, summer 2020 edition. Retrieved November 25, 2021, from https://steamtraen.blogspot.com/2020/06/the-gueguen-saga-update-summer-2020.html

Carpenter, C. J. (2013). A Meta-Analysis of the Effectiveness of the "But You Are Free" Compliance-Gaining Technique. *Communication Studies*, *64*(1), 6–17. https://doi.org/10.1080/10510974.2012.727941

18

Carpenter, C. J., & Boster, F. J. (2009). A meta-analysis of the effectiveness of the disrupt-then-reframe compliance gaining technique. *Communication Reports, 22*(2), 55–62. https://doi.org/10.1080/08934210903092590

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

Channouf, A. (1990). *Antécédents et effets cognitifs et comportementaux des conduites : De l'internalité à la consistance* (These de doctorat). Université Pierre Mendès France (Grenoble). Retrieved November 25, 2021, from https://www.theses.fr/1990GRE29032

Delacre, M., Lakens, D., Ley, C., Liu, L., & Leys, C. (2021). Why hedges' g's based on the non-pooled standard deviation should be reported with welch's t-test. https://doi.org/10.31234/osf.io/tu6mp

DellaVigna, S., & Linos, E. (2020). *RCTs to Scale: Comprehensive Evidence from Two Nudge Units* (Working Paper No. 27594). National Bureau of Economic Research. https://doi.org/10.3386/w27594

Desrumaux, P. (1996). Hypothese de dependance entre les processus d'explications causales internes et externes et l'engagement pro-attitudinal: Une application au don du sang. *Revue internationale de Psychologie sociale, 9*, 77–94.

Dolinska, B., & Dolinski, D. (2006). To command or to ask? Gender and effectiveness of "tough" vs "soft" compliance-gaining strategies. *Social Influence, 1*(1), 48–57. https://doi.org/10.1080/15534510500314571

Dufourcq-Brana, M. (2007). *L'influence d'une déclaration de liberté sur l'efficacité du pied-dans-la-porte et de l'amorçage* (These de doctorat). Lorient. Retrieved November 25, 2021, from http://www.theses.fr/2007LORIS100

Dufourcq-Brana, M., Pascual, A., & Gueguen, N. (2006). Abstract. *Revue internationale de psychologie sociale, 19*(3), 173–187. Retrieved November 25, 2021, from https://www.cairn.info/revue-internationale-de-psychologie-sociale-2006-3-page-173.htm

Duval, S., & Tweedie, R. (2000). A Nonparametric "Trim and Fill" Method of Accounting for Publication Bias in Meta-Analysis. *Journal of the American Statistical Association, 95*(449), 89–98. https://doi.org/10.1080/01621459.2000.10473905

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Farley, S. D., Kelly, J., Singh, S., Jr, C. T., & Young, T. (2019). "Free to Say No": Evoking freedom increased compliance in two field experiments. *The Journal of Social Psychology, 159*(4), 482–489. https://doi.org/10.1080/00224545.2018.1505707

Fillon, A., Kutscher, L., & Feldman, G. (2021). Impact of past behaviour normality: Meta-analysis of exceptionality effect. *Cognition and Emotion, 35*(1), 129–149. https://doi.org/10.1080/02699931.2020.1816910

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in Disguise—An Experimental Study on Cheating. *Journal of the European Economic Association, 11*(3), 525–547. https://doi.org/10.1111/jeea.12014

Garmendia, C. A., Nassar Gorra, L., Rodriguez, A. L., Trepka, M. J., Veledar, E., & Madhivanan, P. (2019). Evaluation of the Inclusion of Studies Identified by the FDA as Having Falsified Data in the Results of Meta-analyses: The Example of the Apixaban Trials. *JAMA Internal Medicine, 179*(4), 582–583. https://doi.org/10.1001/jamainternmed.2018.7661

Gehanno, J.-F., Rollin, L., & Darmoni, S. (2013). Is the coverage of google scholar enough to be used alone for systematic reviews. *BMC Medical Informatics and Decision Making, 13*(1), 7. https://doi.org/10.1186/1472-6947-13-7

Grassini, A., Pascual, A., Guéguen, N., & Jacob, C. (2012). The effect of the 'evoking freedom' technique on sales in a computer-mediated field setting. *The International Review of Retail, Distribution and Consumer Research, 22*(4), 435–437. https://doi.org/10.1080/09593969.2012.690780

Grosch, K., & Rau, H. A. (2016). *Gender differences in compliance: The role of social value orientation* (Working Paper No. 88). GlobalFood Discussion Papers. Retrieved November 25, 2021, from https://www.econstor.eu/handle/10419/146902

Guéguen, N., Jacob, C., Pascual, A., & Morineau, T. (2002). Request Solicitation and Semantic Evocation of Freedom: An Evaluation in a Computer-Mediated Communication Context. *Perceptual and Motor Skills, 95*(1), 208–212. https://doi.org/10.2466/pms.2002.95.1.208

Guéguen, N., Joule, R.-V., Halimi-Falkowicz, S., Pascual, A., Fischer-Lokou, J., & Dufourcq-Brana, M. (2013). I'm free but I'll comply with your request: Generalization and multidimensional effects of the "evoking freedom" technique. *Journal of Applied Social Psychology*, *43*(1). https://doi.org/10.1111/j.1559-1816.2012.00986.x

Guéguen, N., & Pascual, A. (2000). Evocation of freedom and compliance: The but you are free of… technique. *Current Research in Social Psychology*, *5*, 264–270.

Hamamura, T., Bettache, K., & Xu, Y. (2018). Individualism and Collectivism. In *The SAGE Handbook of Personality and Individual Differences: Volume II: Origins of Personality and Individual Differences* (pp. 365–382). SAGE Publications Ltd. https://doi.org/10.4135/9781526470317

Hedges, L. V. (1992). Meta-Analysis. *Journal of Educational Statistics*, *17*(4), 279–296. https://doi.org/10.3102/10769986017004279

Higgins, J., Chandler, T., Cumpston, M., Li, T., Page, M., & Welch, V. ((2022). *Cochrane handbook for systematic reviews of interventions version 6.3 (updated february 2022)*. Cochrane. www.training.cochrane.org/handbook

Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. (2022). *Cochrane handbook for systematic reviews of interventions version 6.3*. Cochrane. www.training.cochrane.org/handbook

Jonas, E., Graupmann, V., Kayser, D. N., Zanna, M., Traut-Mattausch, E., & Frey, D. (2009). Culture, self, and the emergence of reactance: Is there a "universal" freedom? *Journal of Experimental Social Psychology*, *45*(5), 1068–1080. https://doi.org/10.1016/j.jesp.2009.06.005

Kiesler, C. A. (1971). *The psychology of commitment: Experiments linking behavior to belief*. Academic Press.

Kiesler, C. A., & Sakumura, J. (1966). A test of a model for commitment. *Journal of Personality and Social Psychology*, *3*(3), 349–353. https://doi.org/10.1037/h0022943

Kim, H. S., & Sherman, D. K. (2007). "Express yourself": Culture and the effect of self-expression on choice. *Journal of Personality and Social Psychology*, *92*(1), 1–11. https://doi.org/10.1037/0022-3514.92.1.1

Long, D. A., Mueller, J. C., Wyers, R., Khong, V., & Jones, B. (1996). Effects of Gender and Dress on Helping Behavior. *Psychological Reports*, *78*(3), 987–994. https://doi.org/10.2466/pr0.1996.78.3.987

Marchand, M., Halimi-Falkowicz, S., & Joule, R. .-. (2009). Comment aider les résidents d'une maison de retraite à librement décider de participer à une activité sociale ? Toucher, « vous êtes libre de… » et pied-dans-la-porte. *European Review of Applied Psychology*, *59*(2), 153–161. https://doi.org/10.1016/j.erap.2008.05.001

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & Delgado López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, *12*(4), 1160–1177. https://doi.org/10.1016/j.joi.2018.09.002

McGuinness, L. A., & Higgins, J. P. T. (2021). Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Research Synthesis Methods*, *12*(1), 55–61. https://doi.org/10.1002/jrsm.1411

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Motyl, M., Demos, A. P., Carsel, T. S., Hanson, B. E., Melton, Z. J., Mueller, A. B., Prims, J. P., Sun, J., Washburn, A. N., Wong, K. M., Yantis, C., & Skitka, L. J. (2017). The state of social and personality science: Rotten to the core, not so bad, getting better, or getting worse? *Journal of Personality and Social Psychology*, *113*(1), 34–58. https://doi.org/10.1037/pspa0000084

Nuijten, S. E. bibinitperiod M. B. (2018). Statcheck: Extract Statistics from Articles and Recompute p Values. Retrieved November 25, 2021, from https://CRAN.R-project.org/package=statcheck

Olkin, I., Dahabreh, I. J., & Trikalinos, T. A. (2012). GOSH – a graphical display of study heterogeneity. *Research Synthesis Methods*, *3*(3), 214–223. https://doi.org/10.1002/jrsm.1053

Pascual, A. (2002). *Soumission sans pression et technique du "vous êtes libre de. . . "* (These de doctorat). Bordeaux 2. Retrieved November 25, 2021, from http://www.theses.fr/2002BOR21000

Pascual, A., Dufourcq-Brana, M., & Guéguen, N. (2002). L'induction d'un sentiment de liberté au service du pied-dans-la-porte : Une évaluation dans le cadre de la communication par ordinateur. *Unpublished*.

Pascual, A., & Guéguen, N. (2002). La technique du "vous êtes libre de...": Induction d'un senti-

ment de liberté et soumission à une requête ou le paradoxe d'une liberté manipulatrice. [The "you are free of..." technique: Induction of a feeling of freedom and compliance in a request or the paradox of a manipulating freedom.] *Revue Internationale de Psychologie Sociale*, *15*(1), 51–80.

Pascual, A., Oteme, C., Samson, L., Wang, Q., Halimi-Falkowicz, S., Souchet, L., Girandola, F., Guéguen, N., & Joule, R.-V. (2012). Cross-Cultural Investigation of Compliance Without Pressure: The "You Are Free to. . ." Technique in France, Ivory Coast, Romania, Russia, and China. *Cross-Cultural Research*, *46*(4), 394–416. https : / / doi . org / 10 . 1177 / 1069397112450859

Pratkanis, A. R. (Ed.). (2007). *The Science of Social Influence: Advances and Future Progress*. Psychology Press. https : / / doi . org / 10 . 4324 / 9780203818565

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*(2), 597–599. https://doi.org/10. 22237/jmasm/1257035100

Schäfer, T., & Schwarz, M. A. (2019). The Meaningfulness of Effect Sizes in Psychological Research: Differences Between Sub-Disciplines and the Impact of Potential Biases. *Frontiers in Psychology*, *10*, 813. https://doi.org/10.3389/fpsyg. 2019.00813

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving Decisions About Health, Wealth, and Happiness.* (Updated édition). Penguin Books.

Triandis, H. C. (1989). The self and social behavior in differing cultural contexts. *Psychological Review*, *96*(3), 506–520. https : / / doi . org / 10 . 1037 / 0033-295X.96.3.506

Vaughn, A. J., Firmin, M. W., & Hwang, C.-e. (2009). Efficacy of request presentation on compliance. *Social Behavior and Personality: an international journal*, *37*(4), 441–449. https://doi.org/10. 2224/sbp.2009.37.4.441

Walters, W. H. (2007). Google Scholar coverage of a multidisciplinary field. *Information Processing & Management*, *43*(4), 1121–1132. https://doi. org/10.1016/j.ipm.2006.08.006

Yeung, S. K., Fillon, A., Protzko, J., Elsherif, M., Moreau, D., & Feldman, G. (2021). Experimental meta-analysis registered report template. *OSF*. https: //doi.org/10.17605/OSF.IO/YTGRP