Comprehensive Results in Social Psychology

ISSN: 2374-3603 (Print) 2374-3611 (Online) Journal homepage: https://www.tandfonline.com/loi/rrsp20

Are tearful individuals perceived as less competent? Probably not

Janis H. Zickfeld, Niels van de Ven, Thomas W. Schubert & Ad Vingerhoets

To cite this article: Janis H. Zickfeld, Niels van de Ven, Thomas W. Schubert & Ad Vingerhoets (2018) Are tearful individuals perceived as less competent? Probably not, Comprehensive Results in Social Psychology, 3:2, 119-139, DOI: 10.1080/23743603.2018.1514254

To link to this article: https://doi.org/10.1080/23743603.2018.1514254

COMPREHENSIVE RESULTS IN SOCIAL PSYCHOLOGY
2018, VOL. 3, NO. 2, 119–139
https://doi.org/10.1080/23743603.2018.1514254

EASP European Association of Social Psychology
Routledge Taylor & Francis Group

ARTICLE

Are tearful individuals perceived as less competent? Probably not

Janis H. Zickfeld[a,b], Niels van de Ven[c], Thomas W. Schubert[a,d] and Ad Vingerhoets[e]

[a]Department of Psychology, University of Oslo, Oslo, Norway; [b]MZES, University of Mannheim, Mannheim, Germany; [c]Department of Marketing, Tilburg University, Tilburg, the Netherlands; [d]Centro de Investigação e Intervenção Social, Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal; [e]Department of Medical and Clinical Psychology, Tilburg University, Tilburg, the Netherlands

ABSTRACT

What are the social signals of emotional tears? This question has fascinated scholars ever since Darwin. Studies have suggested several interpersonal effects of emotional tears. A recent article presented evidence in three studies that tearful individuals are not only perceived as warmer but also less competent than their non-tearful counterparts. However, the competence effect was relatively small, and a recent replication failed to find such an effect in two different populations while the warmth effect was replicated. This questions the generalizability of the effect of tears on perceived competence. To test whether individuals expressing emotional tears are really perceived as less competent and what boundary conditions such an effect might have, we specified a decision tree of three different studies in which we test differences between the original reference study and the replication. We replicated previous findings of the perceived (in)competence in Study 1 (n = 531) but observed a considerably smaller nonsignificant effect when proceeding to Study 2 (n = 471), which increased the number of stimuli. The earlier and now repeated replication failure can therefore likely be attributed to an increased variation in stimulus material. We conclude that there is not enough evidence to argue that one social outcome of tears signals a relative lack of competence as the effect seems to depend on the specific stimuli used.

ARTICLE HISTORY
Received 24 October 2017
Accepted 16 August 2018

KEYWORDS
Emotional tears; competence; stereotype content model

Tears are a strong social signal, and people try to regulate both their own and others' crying to adapt to social norms and manage the impression they are making. But what exactly do tears convey, and are those signals their function? This question has fascinated researchers throughout the last decades. While Darwin (1872) interestingly attached no purpose to emotional tears, recent studies have explored the intra-individual and inter-individual functions of tears alike (Gračanin, Bylsma, & Vingerhoets, 2017). A body of research investigating the social aspects of tears has provided consistent evidence that tearful individuals are perceived as warm (Van de Ven, Meijs, &

CONTACT Janis H. Zickfeld jhzickfeld@gmail.com
Supplementary material can be accessed https://osf.io/fxnrp/here.

Vingerhoets, 2017; Zickfeld & Schubert, 2018), which triggers an approach tendency in others to provide social support for the crying person (Balsters, Krahmer, Swerts, & Vingerhoets, 2013; Van de Ven et al., 2017; Vingerhoets, Van de Ven, & Van Der Velden, 2016; Zickfeld & Schubert, 2018).

Past work not only found positive effects toward those who cry. For example, people sometimes tend to avoid those who cry (Seidel, Habel, Kirschner, Gur, & Derntl, 2010), think that someone who cries is rather helpless (Vingerhoets et al., 2016), and male individuals are rated as less competent when they express tears in a work-related context in contrast to crying women (Fischer, Eagly, & Oosterwijk, 2013).[1] Based on these prior findings and the stereotype content model (S. T. Fiske, Cuddy, Glick, & Xu, 2002) that suggests that *warmth* and *competence* are the two dimensions on which people are evaluated first, Van de Ven et al. (2017) predicted and found that tearful individuals are not only seen as warmer but also as less competent.

However, a recent replication study found that someone shedding emotional tears was seen as warmer but failed to replicate that it led to a lower perceived competence of the tearful person (Zickfeld & Schubert, 2018). This questions whether the initial observation is real. As the replication attempt contained several small differences compared to the original studies, it is not clear whether the replication failed because the competence effect actually does not exist, or because contextual moderators were present. In a nutshell, the primary goal of the current research is, therefore, to assess whether the findings by Van de Ven et al. (2017) represent an actual effect and if so, what the boundary condition of such a competence effect would be.

Studying the perception of tears is important, as they are displays of an underlying emotional experience and the associated behavioral intentions (Gračanin et al., 2017). The reactions an emotion elicits in other people are a key part of the function of the emotion. Keltner and Haidt (1999) for example argue that emotions signal to others what someone feels, allowing complementary and reciprocal reactions in others. Van Kleef's (2009, 2016) emotion-as-social-information model indicates that seeing another person's emotional display triggers inferential processes about what they feel that can then activate affective responses in observers as well. Consider the case of anger, where the display of anger signals to another person that the angry person is highly dissatisfied with the situation, allowing observers to take this information into account, while at the same time influencing impressions formed of the angry individual (Sell, Cosmides, & Tooby, 2014). Tears are thought to serve such a similar function, for example as signals of sadness and/or helplessness that elicit sympathy and helping behavior (Vingerhoets & Bylsma, 2016). Therefore, finding out why Van de Ven et al. (2017) found different results than Zickfeld and Schubert (2018) did is an essential first step to gain a better understanding of the cognitive, affective, as well as the behavioral responses that tears elicit.

Overview of target studies

Van de Ven et al. (2017) (hereafter *VMV*) found in three studies (total $N = 1042$) that crying individuals were rated as lower on competence than when they rated the same pictures with the tears digitally removed. VMV's studies each had a slightly different study design. The stimuli in each study were derived from a set used in earlier research by Vingerhoets et al. (2016), which consisted of portraits of visitors to the exhibition *The*

Artist is Present by the Serbian performance artist Marina Abramović (Anelli, 2012). Vingerhoets et al. chose 20 pictures of crying individuals, indicated by moist eyes or tears running down their cheeks. They then created control stimuli by removing the tears digitally. Of these 20 pictures, 12 show female and 8 show male targets.

In Study 1 (n = 172), VMV had used 10 of these pictures (5 males and 5 females), each also in a tearful and non-tearful variant, resulting in 20 pictures in total. Dutch undergraduates were then presented with all 20 pictures in a within-subjects design and completed 8 items measuring perceived warmth and competence for each image (S. T. Fiske et al., 2002). In Study 2 (n = 653), US participants were sampled via Amazon MTurk. The authors picked two pictures (one male, one female) as the stimuli, which were again shown in a tearful and non-tearful variant. This time, in a 2 (target gender: male vs. female) × 2 (tears vs. no tears) between-subjects design, participants were randomly allocated to one of the four possible pictures. Afterward, they completed the same measures on perceived competence and warmth, as well as an additional measure of perceived sadness and a question about other emotions that the target was experiencing. In the final Study 3 (n = 217), VMV only employed the female picture used in Study 2. Study 1 and 2 found effects for both male and female targets, which were, if anything, weakest for women. This is why VMV in Study 3 only focused on a female target, as that seemed the more conservative test. Again, they used a between-subjects design with Dutch undergraduate participants being either presented with the tearful or non-tearful picture. In addition to the same measures as in Study 2, participants were now asked to indicate the likelihood that they would approach the person and that they would collaborate with the depicted target. An overview of the study design is provided in Table 1.

Zickfeld and Schubert (2018; hereafter referred to as *ZS*) attempted to replicate the findings of the warmth and competence effect, with (a selection from) the same stimuli. They conducted two studies (total N = 561) with the same research method, one using US MTurk workers (n = 350) and one using Norwegian participants (n = 211). They used the same stimulus pool as VMV but included all 20 different targets in both variants, resulting in 40 stimuli (each target with and without the tear present). Each participant was then exposed to a random subset of six pictures from this total set (always three males and three females). The participants were however only shown tearful or non-tearful pictures, rendering this a between-subjects factor. Participants then completed all measures as in VMV's Study 2. Finally, ZS added the two questions on helping and working intentions from Study 3 of VMV for each of the six presented faces and added an item asking how *touched* and *moved* participants perceived the targets to be.

Basically, ZS combined all of VMV's studies: it followed the between-subjects design and procedure of VMV's Study 2 and 3 but used more than one stimulus (as had been the case in VMS's Study 1, though with fewer stimuli per participant). It included all measures from Study 2 and 3 and added an item on whether people felt moved. The total set of stimuli was larger, and ZS included stimuli that VMV had not. An overview of the actual material and design is provided in Table 1.

ZS replicated that tearful individuals were seen as warmer but did not replicate the finding that they were rated as less competent. Their results suggested the absence of a meaningful effect (Lakens, 2017). In addition, a recent unpublished study using

Table 1. Overview of participant characteristics, study design, number of stimuli, and measures for studies by Van de Ven et al. (2017), Zickfeld and Schubert (2018), and one unpublished study.

Study	Participants	Design	Manipulated factor	Stimuli[a]	Measures (α)
VMV					
Study 1	Dutch undergraduates (M_{age} = 19.68)	2 × 2 Within	Tears (within) Target gender (within)	10 targets (5 male, 5 females)	PW (.70–.93) PC (.70–.93)
Study 2	Amazon MTurkers (M_{age} = 34.33)	2 × 2 Between	Tears (between) Target gender (between)	2 targets (1 male, 1 female)	PW (.93) PC (.91) PS ER
Study 3	Dutch undergraduates (M_{age} = 20.00)	2 Between	Tears (between)	1 female target	PW (.86) PC (.86) IH IW
ZS					
Study 1	Amazon MTurkers (M_{age} = 37.87)	2 × 2 × 3 Mixed	Tears (between) Target gender (within)	20 targets (8 male, 12 female)	PW (US: .95, NO: .93) PC (US: .92, NO: .92) PS PTM ER IH IW
Study 2	Norwegian normal adults (M_{age} = 40.04)	2 × 2 × 3 Mixed	Tears (between) Target gender (within)	20 targets (8 male, 12 female)	Same as ZS Study 1
Unpublished Study	Norwegian undergraduates (M_{age} = 23.40)	2 × 2 × 3 Within	Tears (within) Target gender (within) Context (within)	20 targets (8 male, 12 female)	PW (.95) PC (.91) PS PTM

PS: Perceived sadness; ER: emotion ratings; IH: intention to help; IW: intention to work with; PTM: perceived touched and moved; PW: perceived warmth; PC: perceived competence.

[a]Stimuli are always available as a tearful picture and the exact counterpart with tears digitally removed. So, the total amount of stimuli is therefore always equal the double amount of different pictures. Perceived warmth and competence items were always presented in a fixed order. In VMV Study 1 and 3, warmth items were presented first and then competence items. In VMV Study 2, this order was reversed. In ZS Study 1 and 2, competence items were always presented first.

Table 2. Overview of perceived competence ratings for studies by Van de Ven et al. (2017), Zickfeld and Schubert (2018), an unpublished study, and the present studies including a meta-analytic effect size for all studies.

	No tears		Tears		Cohen's *d* [95% CI]		
Study no.	*n*	*M* (SD)	*n*	*M* (SD)	Male	Female	Overall
VMV							
Study 1	1716	3.70 (1.04)	1712	3.42 (1.06)	−.31 [−.41, −.22]	−.23 [−.32, −.13]	−.27 [−.33, −.20][c]
Study 2	317	3.74 (1.15)	336	3.37 (1.20)	−.47 [−.70, −.24]	−.22 [−.44, −.01]	−.31 [−.47, −.16]
Study 3	114	2.96 (1.10)	103	2.36 (.88)	–	−.60 [−.87, −.32]	−.60 [−.87, −.32]
Overall[a]					−.35 [−.49, −.22]	−.32 [−.54, −.11]	−.35 [−.50, −.19]
ZS							
Study 1 (US)	1068	3.35 (1.26)	1054	3.34 (1.31)	−.02 [−.14, .10]	.00 [−.12, .12]	−.01 [−.09, .08]
Study 2 (NO)	605	3.29 (1.22)	637	3.17 (1.29)	−.03 [−.19, .12]	−.15 [−.31, .01]	−.09 [−.21, .02]
Unpublished	222	4.11 (1.32)	444	4.00 (1.25)	−.11 [−.35, .12]	−.05 [−.28, .17]	−.09 [−.25, .07]
Overall[b]					−.19 [−.35, −.02]	−.19 [−.33, −.05]	−.21 [−.35, −.06]
Current manuscript							
Study 1	240	3.68 (1.19)	291	3.15 (1.31)	−.62 [−.87, −.36]	−.27 [−.51, −.03]	−.43 [−.60, −.25]
Study 2	1343	3.57 (1.26)	1421	3.48 (1.25)	−.14 [−.25, −.03]	.02 [−.08, .12]	−.07 [−.15, .00
Overall[d]					−.23 [−.39, −.07]	−.17 [−.29, −.05]	−.21 [−.34, −.09]

For all studies, ratings are on a response scale from 0 to 6.

VMV refers to Van de Ven et al. (2017), and ZS refers to Zickfeld and Schubert (2018). [a]Meta-analytic effect size of all studies by VMV. [b]Meta-analytic effect size of all studies by VMV, ZS, and one unpublished study together. [c]Effect sizes are calculated based on the "classical" Cohen's *d* (Westfall, 2016) and therefore differ from the effect size reported by VMV. [d]Meta-analytic effect size of all studies by VMV, ZS, one unpublished study, and the present two studies together.

Norwegian undergraduates modeled on the design by ZS failed to detect the competence effect as well (see Table 1, Supplementary Material, and https://osf.io/r9s3h/).

Meta-analyzing all three successful attempts provided by VMV results in an estimated effect size of $d = -.35$ [−.50, −.19] (see Table 2). Adding both studies by ZS and the unpublished study reduces this to an overall effect size of $d = -.21$ [−.35, −.06]. Further examination suggests that the overall effect and the competence effect for both male and female targets are still statistically significant (Table 2). Nevertheless, the non-replication raises the question how robust the effect is and whether there are boundary conditions.

Because ZS made some minor changes to the original study, it is interesting and relevant to determine whether the original findings are robust or whether VMV merely obtained a false positive (see also Brandt et al., 2014). If the findings are robust, it is important to examine which differences in the methodology led to the different results obtained by VMV and ZS to be able to understand the boundary conditions better and consider what the theoretical implications are. Next, we will discuss possible reasons for the replication failure and present the preregistered studies designed to clarify whether seeing tearful individual affects how competent they are perceived to be.

Possible explanations for the null effect in the replication

False positive or false negative

The first explanation of why VMV and ZS had different findings is that the effect of VMV could have been a false positive, and the effect thus simply does not exist. The second explanation is that the effect of ZS was a false negative and that, despite the high power of the ZS' studies, the effect actually exists, but that the studies of ZS just failed to find it.

It might also be that seemingly minor differences in the procedure between the studies of VMV and ZS led to the different outcomes of the studies. We focus on the following three possibilities, namely, the stimuli, the addition of different measures, and the specific samples used.

Differences in use of stimuli

ZS included more variation in their stimuli selection. VMV's Study 2 and 3 used a between-subjects design, where participants only saw one picture. This has the advantage of reducing possible carryover effects of one picture on others but has at the same time the disadvantage that the effect might not generalize beyond the pictures used in those studies.[2] Limiting the study selection in such a way might pose problems to generalizability.

Note that the first study by VMV did include multiple stimuli. It used a within-subjects design, in which participants saw 10 different targets twice, both with and without visible tears, and found the competence effect. One possible explanation for why VMV found an effect and ZS did not is that participants in VMV's Study 1 were watching both tearful and non-tearful faces, which could have led to a contrast effect where perceived differences are exacerbated. In ZS studies, participants were only exposed to either tearful or non-tearful pictures. A second difference between Study 1 of VMV and that of ZS is that ZS used the full set of 20 target persons, while VMV used a selection of 10 target persons in their Study 1.[3]

The addition of an extra measure

ZS also added an item asking whether the target was perceived as being *moved* or *touched* based on the *kama muta* framework (A. P. Fiske, Seibt, & Schubert, 2017). Kama muta is considered to be a positive social emotion that is evoked by sudden intensifications of communal relationships and often (but not always) labeled as being *moved* or *touched*. If intense enough, it is often accompanied by chills, a warm feeling in the chest, and, most importantly, tears. ZS wondered if adding an item asking about being *moved* or *touched* might lead participants to perceive the tearful pictures differently. While tearful individuals might intuitively often be perceived as sad and/or helpless, which can lead to the lower perception of competence (Seidel et al., 2010), providing participants with the more positive explanation for the target's tears (resulting from *being moved*) may actually prevent the detrimental effect on judged competence. However, adding that measure might have also primed people to look for this option, which might not be their initial interpretation of tears (Kühnen, 2010).

Samples

A further difference between both studies is that VMV used Dutch undergraduates and US-based MTurk workers, while ZS sampled Norwegian participants and MTurk workers. Because the results of VMV and ZS were also different in the MTurk samples they both used, it seems that differences in the samples can be ruled out as an explanation for the

replication failure. Furthermore, we also do not see a clear theoretical reason why Norwegian participants are expected to respond differently from Dutch or US participants (nor do we find differences between the samples in how they respond to the other measures such as perceived warmth).

The present studies

In order to explore why ZS did not replicate the competence effect found by VMV, we have joined forces. The principal aim of the present article is to test whether the competence effect as obtained by VMV exists, and if so, what its boundary conditions are. The studies will also continue to evaluate the effect of tears on perceived warmth (an effect both VMV and ZS consistently found). For testing the original effect and ruling out different explanations, we have identified three different stages of testing the effect. The stages are best understood as a decision tree: performing the second experiment depends on the outcome of the first experiment, and so on. The decision tree is displayed in Figure 1. We first briefly describe the main idea of the tests conducted at each stage, after which we provide the study methods for each study in more detail.

Stage I

First, at Stage I, we perform a close replication of VMV's Study 2 with the same stimuli (see Brandt et al., 2014; we use the term *close* replication instead of *direct* or *exact* as no replication can be regarded as completely *direct* or *exact* [e.g. different participants], but rather *as close as* possible to the reference study). We use Study 2 as the basis as it includes both male and female targets, used identical measures as ZS' replication attempt, and revealed the lowest effect size. If we fail to replicate the competence effect there, we conclude that we do not have enough evidence for the effect based on the meta-analytical effect size of all studies and stop testing. The conclusion is then that the effect of tears on perceived competence found by VMV was a false positive and the theory put forward is not supported. Note that this stage does contain a measure of being moved that was not included in the original study of VMV. However, as this question is asked at the end and because there is only one target person being evaluated, this question cannot influence the competence findings.

Stage II

However, if we replicate the original finding in Stage I, we will continue to Stage II. This is a close replication of ZS study, with the exception that we leave out the question on whether people think the person displaying the tear feels moved (the addition of which we leave for Stage III). Stage II, therefore, introduces the first main difference between VMV's and ZS' studies, namely, increasing the number of stimuli (compared to VMV Study 2), without moving to a full within-subjects design that has participants rate the same person with and without tears (as in VMV Study 1). We, therefore, increase the total number of stimuli from 4 (2 targets, each with and without the tear present) to 40 (20 targets, each with and without the tear) as ZS had done. Each

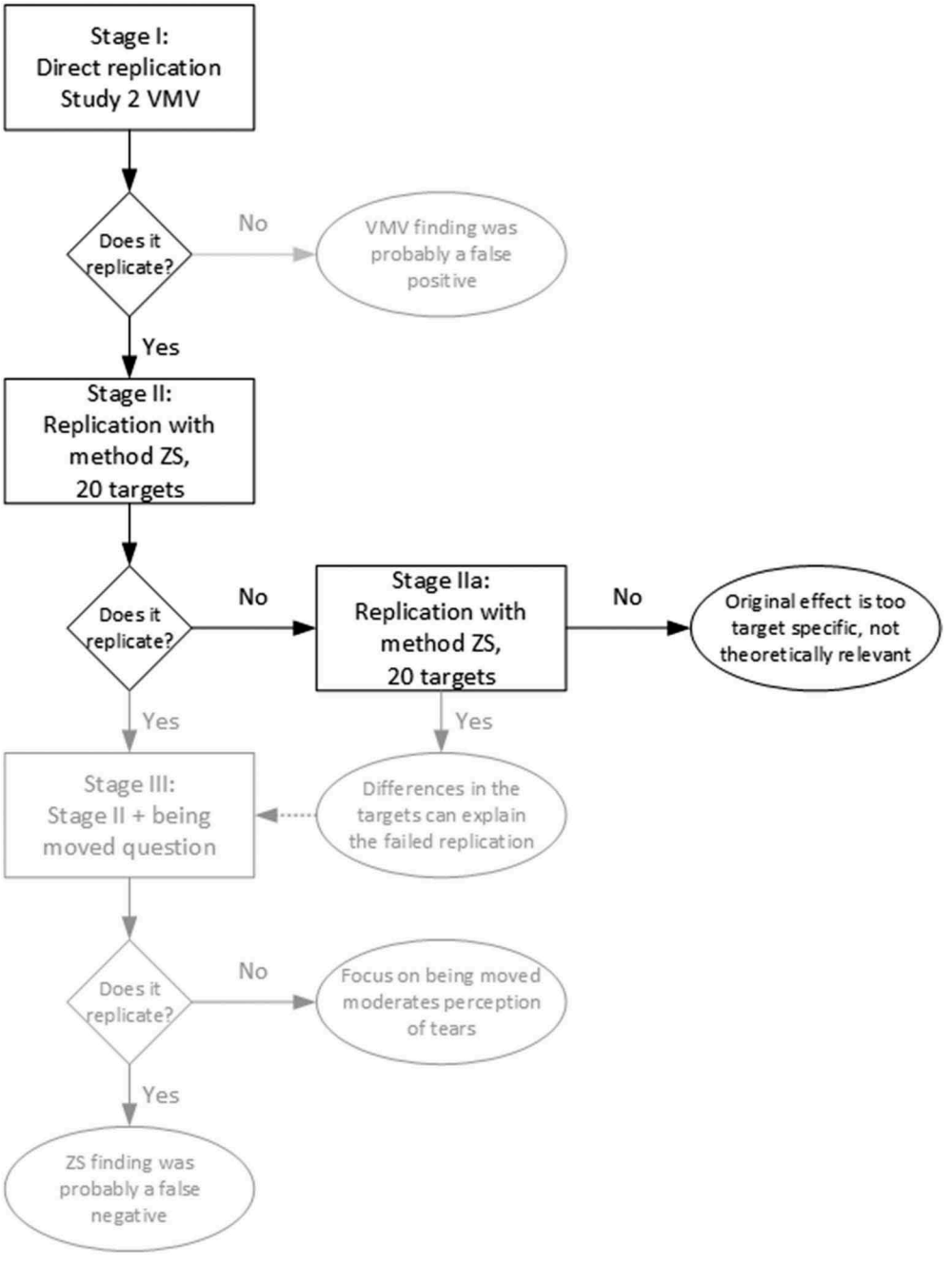

Figure 1. Flowchart of study process.

The steps we took are presented in black, those in grey were the steps we planned but did not do due to the findings. The dotted line indicates that after this conclusion, the setup of Stage III would have changed slightly based on what we learned, as we would only use a selection of the targets.

participant sees six stimuli (all with or without tears) to make the study manageable for the participant (as in ZS). If we find the effect of visible tears on perceived competence, we move to Stage III.

If we do not find the effect at Stage II, we move to Stage IIa and reanalyze the data of the study to test whether there is a difference between the 10 targets that had been used in Study 1 of VMV and the 10 additional targets that ZS had added. If the results show that responses are different such that the perceived competence is affected by the tears of the 10 targets used by VMV, but not by the additional 10 used by ZS, there might be specific differences in how these sets are perceived. Were the tears more visible for the original 10 than the 10 that were added later? Was the warmth effect (that we consistently find) also lower for this newly added set? If so, the conclusion is likely that the tears were not visible enough in the newly added targets, and if these tears are not really visible in this extra set, the competence effect is not likely to appear there (see Figure 1).

If there are no differences between the original 10 targets used in VMV and the 10 that were later added by ZS, and Stage II does not replicate the effect, the effect found by VMV that people perceive those who display a tear as less competent is probably too target specific, and we, therefore, conclude that the effect likely only exists for particular targets but is not generalizable. Consequently, we will discontinue experimentation. The final conclusion then is that the effect VMV found is limited to specific targets and therefore the general theory of VMV (displaying tears makes someone look less competent) is not yet supported. Future studies might find possible moderators that explain how and why some tearful faces are seen as less competent, but until then, we should interpret the results as showing no support for the general idea that a tearful face is seen as less competent.

Stage III

If we do find evidence for the competence effect when we add multiple target faces, we continue to Stage III. At this stage, we include the second main difference between VMV and ZS and add the additional item asking how *moved* or *touched* the target is perceived to be. If we find evidence for the competence effect without the additional item, but not when we include it, we conclude that this item might have made people aware that the person need not necessarily be sad but also could feel moved which thus affects their perception. This is then discussed as a potential moderator of the competence effect.

Finally, if we do find evidence for the competence effect even when including the additional item in Stage III, it will lead to the conclusion that ZS' result was likely a false negative or that other moderators that have not yet been identified had caused the difference in results between VMV and ZS. It would strengthen the support for the main theory of VMV again. For each stage, we perform a meta-analysis, combining the obtained effect sizes with the effects presented by VMV and ZS in order to evaluate the state of the evidence regarding the competence effect.

Our main prediction for all studies is based on the hypotheses formulated by VMV. We expect that tearful individuals are perceived as less competent than non-tearful individuals. We also include additional non-focal tests because they were part of the studies by VMV and ZS. These include the hypothesis that tearful individuals are perceived as warmer in contrast to non-tearful individuals. In addition, the path between tears and perceived warmth was not mediated by perceived sadness (Zickfeld & Schubert, 2018), but by perceived being moved as assessed in the respective condition of Stage 1 and 3, if we reach it.

For all studies, participants were presented with informed consent and could discontinue participation at any point in time. All studies were reviewed by the internal review board of the University of Oslo. All data files including syntax and raw data are made available on our project page (https://osf.io/fxnrp/). For statistically nonsignificant findings, we conducted equivalence tests for the main analyses using the *TOSTER* package in order to provide information on the absence of a meaningful effect (Lakens, 2017).

Study 1: replicating Study 2 by VMV

The first study consisted of a close replication of Study 2 by VMV. We included all original measures and stimuli as used by VMV, including the original wordings and exposure details.

Method

Participants

Participants were recruited on Amazon MTurk and compensated with $.18 for a 1.5-min survey as in VMV. We screened for US-based participants in good MTurk standing (>50 completed assignments with >95% acceptance rate). We planned sample size based on the overall effect[4] of VMV Study 2, which reported a Cohen's *d* of −.316. Aiming at a power of .95 and setting the alpha level to .05 (two-tailed), G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) suggested a total sample size of 524 (262 for each condition). In total, we collected 536 participants. Participants were excluded if their responses contained more than 50% of missing data or if they indicated living in a different country than the United States, leaving a final sample of 531 US American participants (239 males, 279 females, 13 missing) ranging from 18 to 76 in age (M_{age} = 35.96, SD_{age} = 11.91).

Methods and materials

We used the same methods as VMV's Study 2 with small adaptations. Participants were asked to form an impression of a person on a picture, and they were randomly allocated to either the tearful (n = 291) or the non-tearful variant (n = 240). We had two stimuli, one female and one male face; participants were allocated to one of them randomly ($n_{\text{no tears, female}}$ = 117, $n_{\text{no tears, male}}$ = 123; $n_{\text{tears, female}}$ = 152, $n_{\text{tears, male}}$ = 139). The study, therefore, used a 2 (condition) × 2(target gender) between design. Participants completed four items targeting perceived competence (*competent, self-assured, skilled*, and *capable*) and four items targeting perceived warmth (*warm, nice, friendly*, and *sincere*) measured on a scale from 0 (not at all) to 6 (very much so). We always measured competence first, because (a) that was the construct of main interest and (b) that is what VMV Study 2 also used. In contrast to VMV and ZS (Table 1), the order of the four competence items, as well as the order of the four warmth items, was randomized for each participant. Perceived sadness was measured with the item: "How sad does this person appear to you?" on the same scale.

Finally, participants were asked which emotions they thought the depicted person experienced, offering the answer choices *sadness, anger, joy, disgust, fear*, and *surprise*. We added the options that they saw no emotion and an open-ended option. All of these options were dichotomous responses, provided in random order. Participants could

select multiple categories, one, or none. Afterward, we included an item used by ZS asking "How touched and moved does this person appear to you?" on the same 7-point scale as the other items. As this item was not included in the original study by VMV, we placed it at the very end of the questionnaire so that it could not influence our main questions of interest. In the end, participants were requested to provide demographic information including age, gender, and country.

Results

All analyses were performed using *R (Version 3.4.3)*. The alpha level for all analyses was set to .05. We repeated all analyses as performed by VMV Study 2. A mean score was calculated based on the four warmth items (α = .93) and the four competence items (α = .91), which is hereafter referred to as *perceived warmth* and *perceived competence*. Descriptives for the perceived competence ratings are presented in Table 2. Information on all other variables is presented in Supplementary Table 1.

Main analyses

For our main hypothesis, we conducted a two-way ANOVA with condition (*tear* vs. *no tear*), target gender (*male* vs. *female*), and their interaction as factors and *perceived competence* as the dependent variable. As predicted, we observed a main effect of condition, $F(1, 521) = 20.89$, $p < .001$, $\eta^2 = .04$,[5] and also a main effect of target gender, $F(1, 521) = 44.45$, $p < .001$, $\eta^2 = .11$. Tearful targets were rated as less competent than non-tearful targets and male targets were rated as more competent than female targets (Table 2). We did not observe a significant interaction between condition and target gender, $F(1, 521) = 2.66$, $p = .10$, $\eta^2 = .004$.

Additional analyses

We tested the effects of tears on perceived sadness by conducting a two-way ANOVA including *condition* and *target gender*, as well as their interaction as factors. We found a main effect of both condition, $F(1, 527) = 112.16$, $p < .001$, $\eta^2 = .16$, and target gender, $F(1, 527) = 114.58$, $p < .001$, $\eta^2 = .15$, as well as a significant interaction effect, $F(1, 527) = 4.81$, $p = .029$, $\eta^2 = .006$. Tearful individuals were judged higher in sadness (d = .86 [.68, 1.04]), as well as female targets (d = .86 [.68, 1.04]). The small interaction suggested that the effect was relatively more pronounced for the male target than the female target.

In addition, we computed a chi-square test on the relation between the dichotomous sadness item and condition. Participants perceived tearful individuals more often as sad, $\chi^2(1) = 39.49$, $p < .001$, r = .31 [.22, .40], while non-tearful targets were perceived as more neutral, $\chi^2(1) = 55.41$, $p < .001$, r = .72 [.58, .82]. The other categories indicated no significant differences.

To explore whether the effect of tears on perceived competence is mediated by how sad the target is perceived to be, we also conducted a mediation analysis using the *lavaan* package (Rosseel, 2012) and 10,000 bootstrap samples. Condition was the independent variable, perceived competence the dependent variable, and perceived sadness the mediator. For all mediation analyses, variables included in the mediation

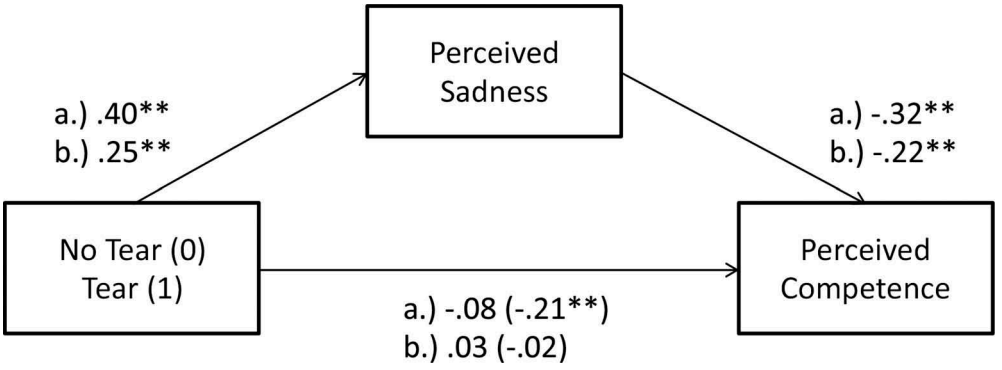

Figure 2. Mediation process of perceived sadness on the effect of tears on perceived competence in Study 1 (a) and 2 (b). Standardized estimates are presented. The original effect without the mediator is presented in parentheses.

*$p < .05$; **$p < .001$.

analysis were centered. We observed a full mediation of sadness on the competence effect, indirect effect $\beta = -.13$ [95% confidence interval (CI) −.17, −.09] (see Figure 2).

In order to test the effect of tears on perceived warmth, we conducted a two-way ANOVA with condition, target gender and their interaction as factors, and *perceived warmth* as the dependent variable. We found a main effect of both condition, $F(1, 523) = 21.20$, $p < .001$, $\eta^2 = .04$, and target gender, $F(1, 523) = 12.70$, $p < .001$, $\eta^2 = .02$, but no significant interaction effect, $F(1, 523) = 1.70$, $p = .192$, $\eta^2 = .003$. Tearful individuals were perceived as warmer ($d = .39$ [.21, .56]), replicating the effect originally found by VMV and replicated by ZS. In addition, female targets were perceived as warmer ($d = .28$ [.11, .45]).

Next, we conducted a mediation analysis using the *lavaan* package with 10,000 bootstrap samples. Condition was employed as the independent variable, perceived warmth as the dependent variable, and perceived sadness as the mediator. As observed in previous studies, sadness did not mediate the warmth effect, $\beta = -.02$ [95% CI −.08, .03] (see Supplementary Figure 3).

Similar to ZS, we tested the effect of tears on perceived being touched and moved by conducting a two-way ANOVA with condition, target gender and their interaction as factors and perceived being touched and moved as the dependent variable. We found a main effect of both condition, $F(1, 524) = 57.99$, $p < .001$, $\eta^2 = .17$, and target gender, $F(1, 524) = 6.02$, $p = .014$, $\eta^2 = .02$, but no significant interaction effect, $F(1, 524) = .23$, $p = .632$, $\eta^2 = .00$. Tearful individuals were perceived as more touched ($d = .91$ [.73, 1.09]), as were female targets ($d = .28$ [.10, .45]).

Subsequently, we conducted a mediation analysis as outlined above using 10,000 bootstrap samples to test whether perceived being touched and moved mediates the effect of tears on perceived warmth as found by ZS. Condition was employed as the independent variable, perceived warmth as the dependent variable, and perceived being touched and moved as the mediator. Perceived being touched fully mediated the warmth effect, $\beta = .16$ [95% CI .11, .21] (see Supplementary Figure 4).

Finally, as done by ZS, we combined both mediation models of perceived sadness and being touched and moved. Conducting a mediation analysis using 10,000 bootstrap

samples, we employed condition as the independent variable, perceived warmth as the dependent variable, and perceived sadness and perceived being touched and moved as mediators. As expected, we observed that the warmth effect was fully mediated by perceived being touched, $\beta = .13$ [95% CI .08, .18], and negatively affected by perceiving the target as sad, $\beta = -.05$ [95% CI −.08, −.02]. The tearful targets increased both ratings of perceived being touched and perceived sadness, but perceived being touched then increased ratings of perceived warmth, while perceived sadness decreased them when controlling for being touched (see Supplementary Figure 5).

Discussion

The first study consisted of a close replication of VMV Study 2 employing the targets used in the original study (one male and one female target). We replicated the findings that tearful targets were perceived as less competent than non-tearful targets ($d = -.24$ [−.38, −.09], overall meta-analytic effect size considering all studies). The effect was comparable to the reference study, though slightly stronger (−.42 [−.59, −.25] vs. −.32 [−.47, −.16]). We also observed stronger differences for the male target than for the female target, as had VMV. In addition, we replicated previous findings that the effect of tears on competence was fully mediated by how sad the targets were perceived. Finally, we found further evidence for the effect that tearful individuals are perceived as warmer and that this effect is mediated by perceiving them as touched and moved, but not due to the perception of sadness, as found by ZS.

Study 2: increasing stimulus variation

As we found an effect of tears on perceived competence in Study 1, we continued with Stage II, which targeted the first main difference between VMV and ZS: the number of stimuli used.

Method

Participants

Participants were again recruited on Amazon MTurk and compensated with $1. Those who participated in Study 1 were excluded from participating using a script (http://uniqueturker.myleott.com). The sample size was calculated based on the effect size obtained by Study 1 ($d = -.43$) and performed using simulations with the *simr* package (Green & MacLeod, 2016). The general model was estimated based on the competence model found in the US sample by ZS. They observed a participant variance of .61, a stimuli variance of .13, and a residual variance of .92. We aimed for a power of at least 95% and set our alpha level at .05. A detailed overview of the calculation procedure is given in the Supplementary Material. Using a Cohen's *d* effect size of −.43 resulted in a total suggested sample size of 240 (120 per condition). In order to retain a power of 95% when only analyzing sets of 10 targets at Stage IIa, we doubled this suggested sample size.

In total, 498 participants were sampled. Participants were excluded if their responses contained more than 50% of missing data, if they indicated that they had already participated in a similar study before, or if they reported living in a different country

than the United States. The final sample consisted of 471 participants (254 males, 202 females, 1 other, 14 missing) ranging from 19 to 71 years of age (M_{age} = 37.42, SD_{age} = 11.78).

Methods and materials

We adapted the procedures and measures from Study 1. Now, participants were shown six different stimuli instead of one. The stimuli were randomly drawn from the pool of 20 targets as used by ZS (see https://osf.io/bmn2v/). Each participant was presented three male and three female targets, randomly chosen. In addition, participants were randomly allocated to either seeing only tearful (n = 245) or only non-tearful stimuli (n = 226). Thus, the present design followed a 2 (between: condition) × 2 (within: target gender) × 3 (within: picture) mixed design similar to the one used by ZS.

After each of the six pictures, participants completed the same measures as in Study 1: perceived competence, perceived warmth, perceived sadness, and the dichotomous measure on the other emotion labels. As explained before, we did not include the measure on whether the target was perceived as *being moved*, to prevent possible carryover effects of that measure on subsequent evaluations of targets. Finally, participants were invited to provide the same demographic information as in Study 1. We also added a question asking whether they had participated in a similar study on Amazon MTurk before.

Results

Again, all analyses were conducted in *R* using the *lme4* package (Bates, 2010). The alpha level for all analyses was set to .05. For all linear mixed models, we allowed intercepts to vary randomly across participants and stimulus type (Judd, Westfall, & Kenny, 2012). We also added all level-1 predictors (target gender) as random slopes. All variables were *z*-standardized on their grand means in order to obtain standardized coefficients (Snijders & Bosker, 2012, p. 53). Again, a mean score was calculated based on the four warmth items (α = .94) and the four competence items (α = .91), which is hereafter referred to as *perceived warmth* and *perceived competence*. Descriptives for the perceived competence ratings are presented in Table 2. Information on all other variables is presented in Supplementary Table 1.

Main analyses

We conducted a linear mixed model with perceived competence as the dependent variable and condition, target gender, as well as their interaction as factors. Main effects for condition, B = −.06, $t(472)$ = −.69, p = .49, and target gender, B = −.03, $t(18)$ = −.15, p = .88, were small and not statistically significant. Perceived competence ratings showed no substantial difference between tearful and non-tearful targets (Table 2). We performed an equivalence test for this difference setting the equivalence bounds to ±.30 for our effect of interest. We found that the observed effect differed significantly from our effects of interest, z = −.09 [90% CI −.17, −.01], p < .001, confirming the absence of an effect we deemed meaningful (Lakens, 2017).

In addition, we did not observe a significant interaction, B = .11, $t(2221)$ = 1.77, p = .08. Although the interaction was not statistically significant, we exploratorily tested

the effect of condition on perceived competence for each target gender separately. We observed a small nonsignificant difference in competence between tearful (M = 3.50, SD = 1.24) and non-tearful women (M = 3.48, SD = 1.26; d = .02 [−.08, .12]), whereas the respective difference between tearful (M = 3.49, SD = 1.27) and non-tearful men (M = 3.67, SD = 1.25) showed a weak effect in the expected direction (d = −.14 [−.25, −.03]).

Stage IIa analyses

As we did not observe a main effect of tears on perceived competence in the main analysis, we repeated the analysis including an additional factor. This factor (*target set*) coded for whether the stimulus referred to the 10 targets used by VMV in their Study 1 or the additional 10 that were also included by ZS. The mixed model, therefore, included condition, target gender, and target set, as well as all two-way and the three-way interaction. In the final model, we did not find a significant interaction between condition and target set, $F(1, 2346) = 3.10$, $p = .08$.[6]

Additional analyses

We again tested whether the path between tears and perceived competence was mediated by perceived sadness. To do so, we computed three different linear mixed regression models. The first regressed the mediator (perceived sadness) on the independent variable (condition) to estimate path *a*. The second regressed the dependent variable (perceived competence) on the mediator and independent variable to estimate path *b* and *c*′. Finally, we estimated path *c* by regressing the dependent on the independent variable. The 95% CI around the indirect effect was estimated using a Monte Carlo simulation (Falk & Biesanz, 2016). We observed a full mediation by perceived sadness, β = −.05, B = −.14 [−.17, −.11] (see Figure 2). Tearful targets were perceived as sadder which in turn reduced the perceived competence of the targets. Although the total effect or the direct effect of the presence of a tear on perceived competence was not significant, the indirect effect was.

We also repeated the same linear mixed model as in the main analysis with perceived sadness as the dependent variable. We found a main effect on condition, B = .95, $t(475) = 10.6$, $p < .001$, but no other significant effects. Tearful individuals were perceived as sadder than their non-tearful counterparts. Perceived warmth was used as the dependent variable in another model. We observed a main effect of condition, B = .37, $t(472) = 4.34$, $p < .001$, and target gender, B = .52, $t(18) = 2.23$, $p = .039$, but no interaction on perceived warmth. Tearful targets and female targets were perceived as warmer. Finally, perceived sadness was also tested as a possible mediator on the relation between condition and perceived warmth using the same procedure as outlined above. We found a small partial mediation of perceived sadness, which was comparable to earlier findings, B = .04 [.01, .07] (Supplementary Figure 3).

Discussion

For the second study, we used the larger pool of targets as done by ZS and observed that the effect on perceived competence decreased drastically (from d = −.43 to d = −.07) and

to non-significance, compared to only using the two specific pictures used in the first study. Using equivalence testing, we got some further indication that the observed effect was smaller than we thought was interesting. As registered, we then compared whether this tiny effect was due to the different stimulus material, but the effect was not particularly stronger when only looking at the targets employed by VMV Study 1. Hence, based on our registered study outline, we did not advance to the third stage but terminated testing after Study 2. Nevertheless, we replicated findings from Study 1 showing that there was an indirect effect of tears on competence via perceived sadness. Finally, we also replicated the consistent finding that tearful individuals were perceived as warmer.

General discussion

Across two studies, we explored differences in findings between two earlier research projects, one showing that tearful individuals are perceived as less competent (Van de Ven et al., 2017) and the other one failing to find similar effects (Zickfeld & Schubert, 2018). We preregistered a decision tree designed to replicate the initial work and to uncover why the replication failed.

In Stage I of the decision tree, we successfully replicated the original Study 2 finding by VMV. We then moved to Stage II, where we added a larger target stimulus pool of tearful persons. With this larger target pool, as it was used by ZS, we no longer found a significant effect of tears on perceived competence. As we no longer found the effect, we conclude that the original effect was too target-specific to be generalizable to an overall conclusion that visible tears have negative effects on perceived competence. We, therefore, did not advance to the final testing stage.

Meta-analyzing the previous studies (VMV, ZS, and an unpublished study) together with the present two studies reveals a small effect of $d = -.21$ [−.34, −.09] (Figure 3). This effect is not smaller than our smallest effect size of interest of $d = \pm.30$ using equivalence testing (TOST, 90% CI [−.32, −.11]). However, it is still questionable whether it is practically important because we observed high heterogeneity of effect sizes in the meta-analysis, $Q(7) = 51.94$, $p < .001$, $I^2 = 90.12$ [75.46, 98.16], which suggests that these effects were affected by a moderating variable. In fact, 95.27% of heterogeneity was accounted for by the fact whether the study employed the full pool of 20 targets in contrast to only 1 target (VMV Study 3), 2 targets (VMV Study 2, this Study 1), or 10 targets (VMV Study 1). These findings challenge the generalizability of a possible competence-diminishing effect of tears. In other words, the effect might work for certain targets but is reduced considerably when more stimulus variation is introduced.

One possible explanation for the reduced effect when increasing stimulus variation could be differences in actual perceived tears. The stimuli employed in the present and previous studies differed in the extent to which tears were visible (see https://osf.io/bmn2v/ for an overview of all targets). Some more subtle manipulations might have failed to produce the expected effect. However, in our current studies (both with the small and large stimulus set), we do replicate the effects VMV found on perceived warmth and sadness, suggesting that the manipulation of the tears was clear enough to affect these variables.

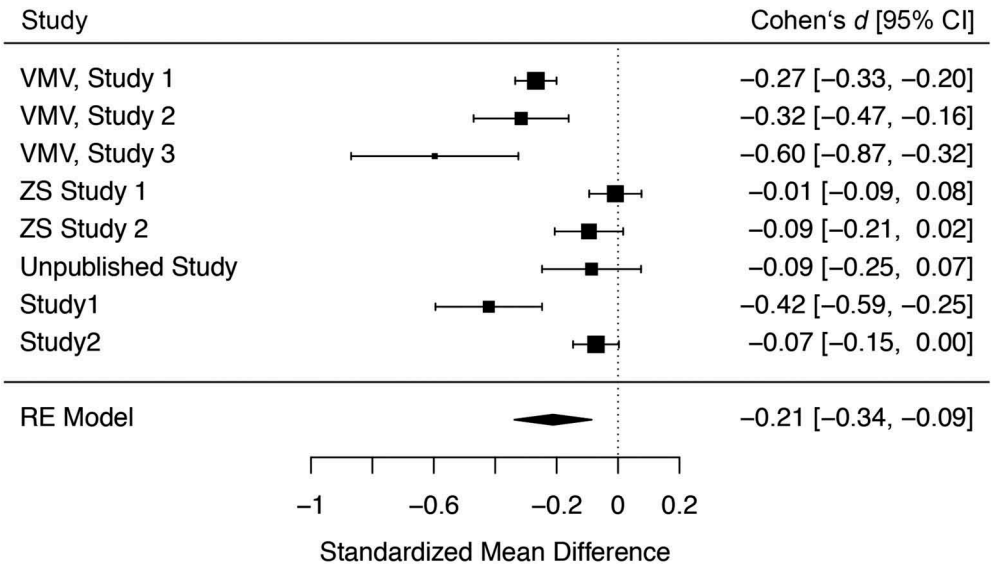

Figure 3. Forest plot of standardized mean differences for perceived competence between non-tearful and tearful targets across VMV, ZS, the unpublished study, and the present study. Heterogeneity tests: $Q(7) = 51.94$, $p < .001$, $I^2 = 90.12$ [75.46, 98.16].

Tearful individuals were perceived as sadder across both studies; we also replicated the previous findings that the perception of sadness was negatively associated with perceiving the targets as competent. In both studies, we replicated an *indirect* effect of tears on perceived competence that was fully mediated by perceived sadness (overall meta-analytical indirect effect across six studies; $\beta = -.07$ [−.11, −.04], Supplementary Figure 7). Thus, it is possible that the small competence effect depends on the degree of perceived sadness. Future studies could directly manipulate the expressed sadness of targets and test whether only targets high in sadness are perceived as less competent.

In fact, the mediation effect might point to an alternative explanation on what the effect of tears is on bystanders. Although VMV hypothesized that tears would increase perceived sadness, and therefore reduce perceived competence, the effect of tears might be much broader. Perhaps, tears serve as a signal that increases the strength of *any* affiliative emotional signal. This fits with the findings of ZS that when someone sheds a tear, this also increases the perception that they are *being moved*. Furthermore, when someone laughs and displays a tear, we might perceive them as having even more joy. As tears are difficult to fake (Vingerhoets, 2013), this could well be a reliable signal that increases the communication function of affiliative emotions.

Another important factor to consider is the difference between male and female targets. Previous research has found that tearful male targets were perceived as less competent than tearful female targets in a work context (Fischer et al., 2013). Notably, this study compared tearful male with tearful female targets but did not include evaluations of non-tearful counterparts. While we did not replicate these findings in the present or previous studies (tearful men were perceived as more or equally competent), we did not provide any context-specific information. Taking into account the findings of Fischer and colleagues, it is possible that a competence effect might be stronger in a particular

context where crying is associated with decreased competence, such as a work environment. Although we did not provide any context-specific information in our studies, we found a slightly stronger overall effect for male targets (d = −.23 [−.39, −.07]) than for female targets (d = −.17 [−.29, −.05]),[7] which was observed throughout most studies except for ZS Study 2. However, these differences were rather small, and they could be related to the fact that we used less male targets in total, thereby possibly introducing less stimulus variation. In addition, most studies did not indicate a significant interaction effect between type of tear and target gender suggesting that there is no systematic evidence for a difference between male and female targets.

A final important factor could be related to the design employed by ZS and for the present Study 2. Participants were always presented with six tearful or six non-tearful pictures in a row, which could have led to a habituation effect. Because the actual randomized order of the targets was not recorded due to a technical problem, we were not able to test this possibility in the present and previous studies (note that this was not registered in our analysis plan). Therefore, we cannot exclude the possibility that a stronger competence effect could be observed for the first cases that were evaluated.

Additional findings

Across both studies, we replicated previous findings by VMV and ZS. First, tearful targets were perceived as warmer than non-tearful targets. Meta-analyzing all effect sizes revealed an overall medium effect of d = .45 [.35, .55] (Supplementary Figure 6). In addition, tearful targets were perceived as sadder across both studies (meta-analytic effect size d = .64 [.50, .77], Supplementary Figure 8) and also as more moved and touched in Study 1 (overall meta-analytic effect size d = .79 [.65, .94], Supplementary Figure 9). Perceiving the targets as being touched and moved fully mediated the fact that tearful individuals were seen as warmer (with a meta-analytic effect size of β = .15 [.13, .17], Supplementary Figure 10), but to a much lower degree by perceived sadness. When controlling for both in the same mediation model, increased perceptions of sadness in fact *reduced* perceptions of warmth. These findings corroborate previous findings, suggesting a link between, on the one hand, being touched and moved and, on the other hand, warmth (Zickfeld & Schubert, 2018; Zickfeld et al., 2018).

Conclusion

The present investigation explored the social signals of tears in depth by testing the proposition whether tearful individuals are perceived as less competent in contrast to non-tearful individuals. Although we replicated the original effect initially, the direct effect of tears on perceived competence was significantly smaller when we used a larger stimuli set. We did consistently find an indirect effect of tears on perceived competence via perceived sadness. Our conclusion is therefore that a general effect of tears on perceived competence is not likely to exist but might emerge only in certain conditions. More broadly, the present findings underscore the need for employing multiple and variable stimulus sets in order to address external validity and generalizability concerns

(see Judd et al., 2012). Based on our and previous studies, tearful individuals are likely to be perceived as warm, but there is no consistent evidence that they are also evaluated as less competent.

Notes

1. Note that in this research, we follow the prior work on crying, testing the perception of tears shed out of sadness or *being moved*. Copying the original studies, we do not add information about the context of tears. Thus, as we study tears that could be perceived as negative, but also positive, we use the term *emotional tears*. Although people might sometimes also shed tears when laughing (Niedenthal, Mermillod, Maringer, & Hess, 2010), this is not what we focus on in this research.
2. Reanalyzing Zickfeld and Schubert's (2018) data only using the two faces employed by Van de Ven et al. (2017) still indicates no significant effect for perceived competence. However, the power of this approach might be too low to detect such an effect (number of cases: United States, n = 188; Norwegian, n = 94).
3. Reanalyzing ZS' data only using the 10 faces employed in Study 1 by VMV still indicates no significant effect for perceived competence in the United States (n = 349) or Norwegian sample (n = 211). However, the power of this analysis might be too low to detect such an effect.
4. Note that although previous findings have pointed at the possibility of an influence of target gender, findings have been inconclusive and generally suggest no difference (see Table 2). As the theoretical basis does not explicitly predict specific gender differences, we therefore rely on the overall effect as our estimate throughout all studies.
5. Please note that we report *eta squared* and **not** *partial eta squared* throughout the manuscript.
6. Nevertheless, we explored the competence effect in the set of 10 targets used by VMV. Again, we observed no significant main and interaction effects. The tearful targets were rated as slightly less competent, B = −.14, $t(459)$ = −1.59, p = .11, and no significant effect of target gender, B = −.19, $t(8)$ = −.83, p = .43. However, we exploratorily repeated the model for each target gender separately. Again, we observed a small effect that tearful male targets were rated as less competent ($M_{\text{no tear}}$ = 3.87, $SD_{\text{no tear}}$ = 1.12 vs. M_{tear} = 3.56, SD_{tear} = 1.23; d = −.26 [−.40, −.12]), while this effect was weaker for female targets ($M_{\text{no tear}}$ = 3.52, $SD_{\text{no tear}}$ = 1.20 vs. M_{tear} = 3.42, SD_{tear} = 1.18; d = −.08 [−.24, .08]).
7. Performing equivalence tests, the effect size for female targets was statistically different from our smallest effect size of interest of d = ±.30 (TOST, 90% CI [−.27, −.07]), while the effect for male targets did not differ (TOST, 90% CI [−.37, −.10]).

Acknowledgment

We are grateful to Marco Anelli for allowing us to use his photographs.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The present research was supported by funding from the Preregistration Challenge by the Center for Open Science.

Author contributions

All the authors conceived and designed the studies. JZ wrote the first draft, and all authors revised it.

ORCID

Janis H. Zickfeld http://orcid.org/0000-0001-7660-2719
Thomas W. Schubert http://orcid.org/0000-0002-3797-4568
Ad Vingerhoets http://orcid.org/0000-0001-8840-6379

References

Anelli, M. (2012). *Portraits in the presence of Marina Abramovic*. Bologna, Italy: Damiani.

Balsters, M. J., Krahmer, E. J., Swerts, M. G., & Vingerhoets, A. J. (2013). Emotional tears facilitate the recognition of sadness and the perceived need for social support. *Evolutionary Psychology*, *11*(1), 147470491301100130.

Bates, D. M. (2010). *lme4: Mixed-effects modeling with R*. Retrieved from http://lme4.r-forge.r-project.org/book/

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*, 217–224.

Darwin, C. (1872). *The expression of the emotions in man and animals*. London, UK: John Murray.

Falk, C. F., & Biesanz, J. C. (2016). Two cross-platform programs for inferences and interval estimation about indirect effects in mediational models. *SAGE Open*, *6*(1), 2158244015625445.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Fischer, A. H., Eagly, A. H., & Oosterwijk, S. (2013). The meaning of tears: Which sex seems emotional depends on the social context. *European Journal of Social Psychology*, *43*(6), 505–515.

Fiske, A. P., Seibt, B., & Schubert, T. W. (2017). The sudden devotion emotion: Kama muta and the cultural practices whose function is to evoke it. *Emotion Review*, 175407391772316. doi:10.1177/1754073917723167

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878.

Gračanin, A., Bylsma, L. M., & Vingerhoets, A. J. J. M. (2017). The Communicative and Social Functions of Human Crying. In J.-M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 217–233). New York, NY: Oxford University Press.

Green, P, & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493-498.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69.

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion*, *13*(5), 505–521.

Kühnen, U. (2010). manipulation checks as manipulation: Another look at the ease-of-retrieval heuristic. *Personality and Social Psychology Bulletin*, *36*(1), 47–58.

Lakens, D. (2017). equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362.

Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The simulation of smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavior and Brain Sciences*, *33*(6), 433–480.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 2.

Seidel, E.-M., Habel, U., Kirschner, M., Gur, R. C., & Derntl, B. (2010). The impact of facial emotional expressions on behavioral tendencies in women and men. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(2), 500.

Sell, A., Cosmides, L., & Tooby, J. (2014). The human anger face evolved to enhance cues of strength. *Evolution and Human Behavior*, *35*(5), 425–429.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks, CA: Sage.

Van de Ven, N., Meijs, M. H. J., & Vingerhoets, A. (2017). What emotional tears convey: Tearful individuals are seen as warmer, but also as less competent. *British Journal of Social Psychology*, *56*(1), 146–160.

Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current Directions in Psychological Science*, *18*(3), 184–188.

Van Kleef, G. A. (2016). *The interpersonal dynamics of emotion*. Cambridge, UK: Cambridge University Press.

Vingerhoets, A. J. J. M. (2013). *Why only humans weep: Unravelling the mysteries of tears*. Oxford, UK: Oxford University Press.

Vingerhoets, A. J. J. M., & Bylsma, L. (2016). The riddle of human crying: A challenge for emotion researchers. *Emotion Review*, *8*(3), 207–217.

Vingerhoets, A. J. J. M., Van de Ven, N., & Van Der Velden, Y. (2016). The social impact of emotional tears. *Motivation and Emotion*, *40*(3), 455–463.

Westfall, J. (2016, March 25). Five different "Cohen's d" statistics for within-subject designs. Retrieved October 3, 2017, from http://jakewestfall.org/blog/index.php/2016/03/25/five-different-cohens-d-statistics-for-within-subject-designs/

Zickfeld, J. H., & Schubert, T. W. (2018). Warm and touching tears: Tearful individuals are perceived as warmer because we assume they feel moved and touched. *Cognition & Emotion*, 1–9. doi:10.1080/02699931.2018.1430556

Zickfeld, J. H., Schubert, T. W., Seibt, B., Blomster, J. K., Arriaga, P., Basabe, N., … Fiske, A. P. (2018). Kama muta: Conceptualizing and measuring the experience often labeled being moved across 19 countries and 15 languages. *Emotion*. doi:10.1037/emo0000450