Journal of Experimental Social Psychology 117 (2025) 104697

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

It is not only *whether* I approach but also *why* I approach: A registered report on the role of action framing in approach/avoidance training effects☆,☆☆

Marine Rougier [a,b,*], Mathias Schmitz [a,c], Ivane Nuel [d], Marie-Pierre Fayant [e], Baptiste Subra [f], Theodore Alexopoulos [f], Vincent Yzerbyt [a]

[a] *Institut de Recherches en Sciences Psychologiques, Université catholique de Louvain, Louvain-la-Neuve, Belgium*
[b] *Department of Experimental-Clinical and Health Psychology, Ghent University, Ghent, Belgium*
[c] *Pontificia Universidad Católica del Perú, Lima, Perù*
[d] *Centre de Recherches sur la Cognition et l'Apprentissage, University of Poitiers, Poitiers, France*
[e] *Laboratoire de Psychologie Sociale, Université Paris Cité, Boulogne-Billancourt, France*
[f] *Laboratoire de Psychologie, EA 4139, University of Bordeaux, Bordeaux, France*

ARTICLE INFO

Keywords:
Framing
Aggression
Approach/avoidance training
Face representation
Reverse correlation

ABSTRACT

Research on approach/avoidance training (AAT) effects shows that approach (i.e., reducing the distance between the self and a stimulus) leads to more positive evaluations of stimuli than avoidance (i.e., increasing the distance between the self and a stimulus). The present experiments relied on a grounded cognition approach to extend this finding by investigating the framing-dependency of AAT effects on facial representations of target stimuli. In a Preliminary Experiment, using antagonistic types of approach (affiliative vs. aggressive) and a reverse correlation paradigm, we found that approach led to more positive facial representations than avoidance when approach was portrayed as affiliative, but this effect decreased and tended to reverse (i.e., yielding more negative facial representations) when approach was portrayed as aggressive. Two registered experiments extended these results while also addressing important limitations of the Preliminary Experiment. First, to prevent any contrast emerging from the joint use of approach and avoidance, Experiment 1 isolated the unique effects of affiliative approach, aggressive approach, and avoidance compared to a control action. We also explored whether aggressive approach and avoidance (two negatively valenced yet distinct actions) produced negative effects characterized by divergent outcomes on facial features (e.g., weak vs. dominant). Second, Experiment 2 tested the importance of the experiential component of approach/avoidance actions by comparing the AAT with a mere instructions condition. Results of Experiments 1 and 2 proved consistent with a framing-dependency of AAT effects. Unveiling the framing-dependency of AAT effects challenges some of the current theoretical views on AAT effects.

Approach and avoidance are among the most fundamental and adaptive actions toward the environment. While positive/negative evaluations facilitate approach/avoidance responses (i.e., a distance reduction/increase between the self and a stimulus, respectively; Solarz, 1960), the reverse relationship also holds. As revealed in the "Approach/ Avoidance Training" (AAT) effect, the repeated performance of approach behaviors yields positive evaluations toward the stimulus, while repetitively executing avoidance behaviors yields negative

☆ This paper has been recommended for acceptance by Rachael Jack.☆☆ The materials, data, and analytic (R) scripts are made publicly available at https://osf.io/q5f9r/?view_only=073b222de7a54b008093f1682cd6511b.This research project started during Marine Rougier's postdoctoral research contract under the supervision of Vincent Yzerbyt. It was supported by Marine Rougier's and Mathias Schmitz's grants from the Fonds de la Recherche Scientifique (FRS-FNRS; grants 1.B347.19 and 1.A393.17, respectively). This project was also supported by a pre-registered research grant from the European Association of Social Psychology. Finally, a part of this research project took place during Marine Rougier's postdoctoral research contract in Ghent University, made possible by the Special Research Fund (BOF) of Ghent University to Jan De Houwer (reference number: BOF22/MET_V/002).

* Corresponding author at: Institut de Recherches en Sciences Psychologiques, Place du Cardinal Mercier, 10 bte L3.05.01, B-1348 Louvain-la-Neuve, Belgium.
E-mail address: m.rougier@hotmail.fr (M. Rougier).

https://doi.org/10.1016/j.jesp.2024.104697
Received 10 May 2022; Received in revised form 4 November 2024; Accepted 7 November 2024
Available online 12 December 2024

evaluations (e.g., Kawakami et al., 2007). So far, AAT effects have been considered without any further specification of the meaning of approach/avoidance. Indeed, approach seemingly evokes positivity, whereas avoidance presumably suggests negativity. However, approach behaviors can sometimes convey a radically different meaning, one akin to negativity instead (e.g., affiliation vs. aggression). In this contribution, our main goal is to test the moderation of the AAT effects by the meaning attached to approach and avoidance actions, that is, the framing-dependency of AAT effects. Relying on a grounded cognition framework, we also consider insights into potential underlying processes.

1.1. The "systematic" outcome of approach/avoidance training effects

The AAT effect is a robust phenomenon. It emerges across a large variety of stimulus types (e.g., novel, social-related, addiction-related) that repeatedly approaching (avoiding) one category of stimuli leads to a more positive (negative) evaluation of this category (Becker et al., 2015; Jones et al., 2013; Rinck et al., 2013; Wiers et al., 2010; Woud et al., 2013). Interestingly, AAT consequences go beyond mere evaluative changes. They also influence behaviors (e.g., sitting distance) toward the category one has been trained to approach (Kawakami et al., 2007), as well as its visual representation in memory (Rougier et al., 2021).

In previous research, two features of the AAT effect stand out, forming what we call a "systematic" AAT outcome. First, the direction of this effect is highly consistent: Approach leads to more positive evaluations (Cacioppo et al., 1993), more positive representations (Rougier et al., 2021) or more favorable behaviors (Kawakami et al., 2007) than avoidance (but see Mertens et al., 2018). Second, this outcome emerges without any further characterization of the meaning of approach and avoidance actions (i.e., without specifying what approach and avoidance stand for in the setting, e.g., Van Dessel et al., 2015; but see Laham et al., 2014).

The standard interpretation of the AAT effect – adopted by various theoretical accounts – is that it results from the evaluative properties of approach and avoidance actions (e.g., Eder & Klauer, 2009; Neumann et al., 2003; Phills et al., 2011; Van Dessel, Hughes, & De Houwer, 2018). Building on the premise that approach/avoidance actions are unequivocally linked to positive vs. negative evaluative features, pairing these actions with stimuli should produce a corresponding change in their evaluation. However, if this very general explanation provides strong grounds for a systematic outcome of AAT effects, it leaves out an important feature of approach and avoidance behaviors, namely, their sensorimotor aspects.

1.2. The grounded perspective on the AAT effect

The AAT effect builds on the seminal idea that the sensorimotor information of approach and avoidance has an impact on evaluations of neutral stimuli (Cacioppo et al., 1993) and echoes the principles of grounded cognition (Barsalou, 1999, 2008; Damasio, 1989; Versace et al., 2014). This view proposes that behaviors such as approach and avoidance are coded in memory as "multimodal representations" with both a) the sensorimotor aspects of approach/avoidance (e.g., movements toward or away from the stimulus, as well as the corresponding visual changes resulting from these movements) and b) the aspects of stimuli that routinely triggered those actions in the past (e.g., the valence of the stimulus or its appearance; Rougier et al., 2018).

Once formed, these multimodal representations are thought to be reactivated during the AAT. Specifically, the sensory activations involved in a typical AAT situation (e.g., when performing approach/avoidance toward groups of faces in the training; Rougier et al., 2021) are assumed to activate the multimodal representations of past approach and avoidance, which includes the visual appearance of previously approached/avoided stimuli (e.g., faces). The simultaneous activations stemming from the present (i.e., perceptions activated from the AAT) and past situations (i.e., perceptions from approach/avoidance representations) would then combine and be reinjected into the memory system (Versace et al., 2014). The new combined information would thus involve the visual features (e.g., facial features) as well as the valence of stimuli one has approached/avoided in the past. Consequently, this approach predicts that approach and avoidance should impact perceived valence of stimuli, as well as their visual representation.

In line with this reasoning, Rougier et al. (2021) showed that approaching vs. avoiding neutral faces in an AAT biased the visual representations of approached vs. avoided groups of faces on a series of facial traits. Specifically, the visual representation of the approached group was evaluated (by independent judges) more positively than the one of the avoided group. Crucially, two results emerged as particularly consistent with a grounded cognition approach. First, the difference in evaluation between visual representations was stronger for facial traits that are most likely to trigger approach/avoidance reactions ("other-relevant" traits, e.g., trustworthy, cruel) than for facial traits less likely to trigger these actions ("possessor-relevant" traits, e.g., intelligent, lazy; see also Wentura et al., 2000). Second, the AAT effect emerged without explicitly mentioning the terms "approach" and "avoidance" in the instructions – that is, when capitalizing on the mere sensorimotor information of approach/avoidance (i.e., the visual impression of walking forward/backward; see also Kawakami et al., 2007; Phills et al., 2011).

1.3. Toward a moderation of AAT by action framing

According to a grounded cognition framework, different types of actions (e.g., affiliative approach vs. aggressive approach vs. avoidance) should be coded as distinct multimodal representations. These representations should contain discrete information about stimuli that were approached/avoided (e.g., affiliative approach could be associated with friendly faces, whereas aggressive approach with unfriendly faces). Indeed, every stimulus that triggered approach or avoidance actions possesses features that are uniquely related to the specific action (e.g., Bossuyt et al., 2014; Krieglmeyer & Deutsch, 2013). Hence, performing a specific action in an AAT (e.g., approach) should evoke distinct multimodal representations (e.g., multimodal representation of affiliative vs. aggressive approach) as a function of the way the action is framed (e.g., affiliative or aggressive). This implies that the framing of the action should moderate the AAT effect.

Some preliminary research supports this idea. For instance, an elaborated contextual framing seems to facilitate the emergence of AAT effects. Indeed, Laham et al. (2014) showed that AAT effects are larger when an elaborated framing contextualizes the approach/avoidance actions (e.g., collecting/rejecting food on an alien planet) compared to when this framing is absent (e.g., pulling/pushing instruction). However, this experiment did not manipulate the meaning of approach/avoidance per se. Other efforts suggest that the direction of the AAT effect can be altered by the evaluative meaning of concurrently approached/avoided stimuli. For instance, Mertens et al. (2018) had participants perform the AAT task on neutral stimuli and conditioned positive vs. negative stimuli at the same time. The authors concluded that the pairing of the approach/avoidance action with a positive or negative stimulus moderates the AAT effect. Yet, this procedure does not allow concluding whether it is the pairing of the movement with a specific valence or the mere valence of the stimulus that is responsible for the effect (for evidence on the effect of stimuli pre-training evaluation, see Krishna & Eder, 2019).[1]

The question of the potential influence of the approach/avoidance meaning is crucial given the context-dependent nature of approach/

[1] We thank an anonymous reviewer for pointing this out.

avoidance behaviors in everyday life. For instance, people can manifest approach for very different reasons (e.g., to hug or to punch someone), resulting in considerable variations on the meaning of this action. Crucially, research on approach/avoidance reactions showed that the activation of these tendencies depends on the way the action is framed in the context of the experiment. For instance, Krieglmeyer and Deutsch (2013) found that participants were faster to approach than to avoid angry faces when approach was framed as aggressive, whereas they were faster to avoid than to approach angry faces when approach was framed as affiliative. These findings suggest that reactions toward the very same stimulus can be diametrically opposed (i.e., tilting from avoidance to approach) as a function of action framing (see also Bossuyt et al., 2014). In the present work, we aim to show that the action's concomitant meaning also plays a role in AAT effects. Specifically, we predict that the action framing should moderate the AAT effect on the visual representation of the stimuli.

1.4. Distinguishing between a grounded cognition approach and alternative models

Generic models of AAT effects stand out as alternative models to the grounded cognition approach and have in common to rely on the evaluative properties of approach (positive) and avoidance (negative) actions. The potential moderation by the action framing, however, fits better with some accounts than with others.

On the one hand, the motivational-systems and self-anchoring accounts would hardly predict a moderation by action framing. The former account states that the evaluative properties of approach and avoidance actions emerge from two core systems of approach and avoidance, with the approach (avoidance) system being generally associated with positive (negative) consequences (Neumann et al., 2003; Wiers et al., 2011). Similarly, according to the self-anchoring account, when approaching a stimulus, the valence of the self-concept, which is usually endowed with a positive valence, is transferred to the stimulus (Kawakami et al., 2007; Phills et al., 2011)[2] – although this association may be more or less favorable as a function of self-esteem (Gawronski et al., 2007). Hence, these two theoretical approaches cannot easily accommodate the flexibility of the link between approach/avoidance and valence, at least when experimentally manipulated by action framing. Therefore, a moderation by action framing should provide evidence against the motivational system and the anchoring accounts.

On the other hand, two other alternative accounts relying on the evaluative properties of approach/avoidance are more compatible with a moderation by the action framing. The common-coding account (Eder & Klauer, 2009) proposes that the valence of approach and avoidance stems from the affective codes associated with the representation of the movement (Van Dessel, Eder, & Hughes, 2018). In the same vein, the inferential account argues that evaluative inferences about stimuli (e.g., "Stimulus A is positive") rest on all activated propositions associated with these actions and how they relate to valence (e.g., "I approached stimulus A", "I generally approach positive things"; Van Dessel, Hughes, & De Houwer, 2018). Although these two approaches rely on different processes (e.g., associative vs. propositional), they both accommodate the idea that the valence of approach/avoidance may depend on framing. Therefore, the moderation by framing effect cannot distinguish the common coding and the inferential accounts from the grounded cognition approach.

Beyond the moderating role of action framing, the role of the experiential component of the AAT (i.e., the effect of actually approaching and avoiding stimuli during the training phase) should enable us to contrast a grounded cognition approach with the inferential account. According to a grounded cognition approach, the enactment of approach/avoidance actions should reactivate more strongly their representations in memory (via sensorimotor simulation) and thus foster AAT effects as compared to merely receiving approach/avoidance instructions. The motivational, self-anchoring, and common-coding accounts would also predict that repeated actions should lead to a larger effect. For these three accounts, repeatedly enacting approach/avoidance should reinforce the associative links as compared to only reading an instruction to approach/avoid. For the inferential account, however, the AAT effect is assumed to be driven by the inferences derived from the instructions. Indeed, previous research showed that merely instructing participants that they will have to approach or avoid novel groups is sufficient to produce an evaluative bias (e.g., Van Dessel et al., 2015; see also Smith et al., 2019). Based on this idea, one can argue that because performing vs. not performing the training involves the same instructions, similar inferences should be activated in both conditions and therefore the experience of approach/avoidance training should not increase the AAT effect.[3]

Interestingly, previous research examined the unique contribution of experiencing approach/avoidance training on the AAT effect by comparing an AAT with mere instructions without performing the training. Results are mixed: Some work showed that experiencing a training has an effect that is independent from instructions (Rougier et al., 2021; Van Dessel et al., 2020, 2016), whereas other work failed to observe such a difference (Smith et al., 2019; Van Dessel et al., 2016). Given the theoretical importance of this question and the mixed results reported in the literature, it is necessary to provide an additional test of the unique role of training as well as a better estimation of its effect size.

Testing the unique contribution of the approach/avoidance training effect also enables to rule out an alternative explanation in terms of demand effects (Corneille & Béna, 2023). One could argue that the instructions provide a simple account for the moderation of the AAT effect by action framing. That is, framing approach as affiliative or as aggressive may come across as an experimental demand to rate approached faces positively or negatively according to the condition. If such an explanation were true, providing instructions alone should be sufficient to generate the predicted moderation effect of action framing, regardless of whether participants do or do not undergo the training. In contrast, if a larger effect emerges in the experience-based procedure, thus showing that experiencing training has a unique effect, this implies that instructions alone (and thus a potential demand effect) cannot fully account for the moderation by action framing (Orne, 2009).

1.5. Overview

The main goal of this paper is to test a grounded cognition approach of the AAT effect. To do so, we investigated the moderating role of action framing on AAT effects, specifically on the visual representation of approached/avoided facial stimuli. In addition, we also examined the importance of actually performing approach/avoidance actions in comparison to merely receiving approach/avoidance instructions.

We report three experiments: A non-registered Preliminary Experiment and two registered experiments (Exp. 1 and 2). The preliminary experiment manipulated antagonistic affiliative versus aggressive approach framings. We relied on Rougier et al.'s (2021) procedure using the Visual Approach/Avoidance by the Self Task (VAAST; Rougier et al., 2018) and the reverse correlation paradigm (Dotsch & Todorov, 2012). We selected these tasks because they usually lead to large AAT effects when combined (Rougier et al., 2021).

In Experiment 1, we aimed at replicating the framing moderation observed in the Preliminary Experiment but with a between-participant

[2] According to the self-anchoring account, the valence does not come from approach/avoidance actions per se but from the self.

[3] Relying on a more recent, qualified inferential view, however, one could argue that experiencing approach/avoidance actions should lead to larger evaluative change (Rougier et al., *in preparation*). We will come back to this point in the General Discussion.

manipulation of the target action (i.e., affiliative approach vs. aggressive approach vs. avoidance; each target action compared to a control action). Again, to the extent that the AAT is framing-dependent, the visual representation obtained in the aggressive approach should result in more negative evaluations than the visual representation in the affiliative approach condition. Moreover, relying on grounded cognition principles, we formulated predictions not only in terms of valence but also with respect to the type of facial features that would yield biased representations (for a similar rationale, see Rougier et al., 2021, Experiments 1–2). We reasoned that if the AAT effect depends on the meaning of the action, the visual representations in the aggressive approach and avoidance conditions should be mostly biased on physical traits typically related to each of these actions.

In Experiment 2, we tested the importance of actually performing approach/avoidance actions in comparison to merely receiving approach/avoidance instructions. If the experiential aspect of approach/avoidance actions contributes to the AAT effect, the effect should be larger when participants enact these actions.

As in Rougier et al. (2021), we had no reasons to predict different effects of approach/avoidance training and approach framing as a function of the characteristics of our samples (e.g., gender, socioeconomic variables, etc.). Nevertheless, face-based social perception can vary based on individuals' cultural environment (e.g., Western vs. Chinese; Sutherland et al., 2018; Wang et al., 2019) or other individual-level variables such as gender or personality (Mattarozzi et al., 2015). For instance, women tend to perceive trustworthy-looking faces as more trustworthy than men and individuals low on agreeability and high on aggressiveness tend to perceive faces as less trustworthy (Mattarozzi et al., 2015). Hence, although we did not vary the cultural background of our sample, we aimed at replicating the framing effect in a diversified sample to assert that this effect could emerge despite potentially increased inter-individual variability in face-based perception. With this, our aim was to increase the generalizability of our findings – that tis, that the framing effect would replicate in a new sample.

We ran the Preliminary Experiment, Experiment 1 (Parts 1 and 2) and Experiment 2 (Part 2) on a crowdsourcing platform (Prolific Academic; www.prolific.co). Samples included native English-speaking US participants (cf. pre-screening criteria) and were heterogeneous, as participants potentially varied greatly in age, gender, religion, ethnic and racial background, or social class, etc. (Gleibs, 2017; Peer et al., 2017). At the same time, Experiment 2 (Part 1) relied on French-speaking undergraduate students from a Belgian university. Students are typically homogeneous and W.E.I.R.D. samples (Henrich et al., 2010). All reported experiments were conducted in accordance with ethical criteria of the American Psychological Association and have been approved by our local ethic committee (2018-20). For each experiment, we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures.

2. Preliminary experiment: initial evidence for the framing-dependency of AAT effects

The aim of this Preliminary Experiment was to gather initial evidence on whether instruction-induced framing moderates AAT effects (for more information on the material or data analyses, see OSF project: https://osf.io/q5f9r/?view_only=073b222de7a54b008093f1682cd6511b). We asked a first sample of participants to repeatedly approach (vs. avoid) neutral faces belonging to novel groups. We additionally manipulated the goal associated with approach actions. Specifically, the instructions informed participants that approach serves affiliation (affiliative condition) or that it serves aggression (aggressive condition). In the affiliative condition, we expected the visual representations (also known as classification images) of the approached stimuli to be rated – by independent judges – as more positive than the avoided ones (i.e., the systematic AAT finding). This difference should decrease or even reverse in the aggressive condition, yielding a significant interaction between movement and action framing. To probe for differences in visual representations, we relied on different (i.e., more or less conservative) strategies divided into two separate parts and involving separate samples (Part 2A with condition-level visual representations and Part 2B with subgroup-level visual representations).

2.1. Power analysis and sample size

As for Experiments 1 and 2, our Preliminary Experiment required two power analyses: One for the participants of the Part 1 (the "face producers" performing the AAT) and one for the participants of Parts 2A and 2B (the "judges"). Of note, chances to detect an effect depends on both parts. The higher the number of face producers in the first sample, the better the quality of the classification images (i.e., the signal/noise ratio) and thus their decipherability by the judges. At the same time, the higher the number of judges, the higher the statistical power to detect any differences between the classification images (see Cone et al., 2021).

Sample size was determined a priori via G*Power (Version 3.1.9.6; Faul et al., 2007; screenshots are available on the OSF repository). We based our face producers' sample size on Rougier et al. (2021, Exp. 1; $N = 93$ per condition) and opted for $N = 150$ participants because of budget constraints (i.e., $N = 75$ per condition). Regarding the judges in Part 2A (i.e., evaluating visual representations at the level of the condition), we relied on the effect sizes of the AAT effect on visual representations obtained from the most similar experiment (i.e., Rougier et al., 2021, Exp. 1; $dz = 1.27$), as well as on the smallest framing-instruction moderation effect obtained in Krieglmeyer and Deutsch (2013; $\eta_p^2 = .09$, i.e., approximately $dz = .63$). With $N = 100$ judges, we had a power of 99.99 % to detect both a main AAT effect and an AAT by framing interaction (5 % false-positive rate in a two-tailed *t*-test with paired samples). This sample afforded 80 % power to detect an effect size of $dz = 0.33$ with a 5 % false-positive rate in a two-tailed *t*-test with paired samples. Regarding the judges in Part 2B (i.e., evaluating visual representations at the level of subgroups), we relied on the effect size of the framing moderation obtained in Part 1 ($N = 101$; $dz = 1.61$). As each participant rated a subset of subgroup-level classification images (i.e., 40 over 256), we increased the total sample of participants to achieve a reasonable number of ratings per image. A sample of 150 participants afforded 99.99 % power to detect an effect of $dz = 1.61$ (5 % false-positive rate in a two-tailed *t*-test with paired samples).

2.2. Method

2.2.1. Part 1: creation of classification images resulting from approach vs. avoidance actions

Participants and Design. One hundred and forty-eight Prolific Academic users ($M_{age} = 37.27$ years, $SD_{age} = 12.36$; 74 males and 74 women; https://prolific.ac/) took part in exchange for a monetary compensation (£5.00/h). We recruited only American English speakers equipped with a computer and with an approval rate of 95 % or more (i.e., to improve data quality, Peer et al., 2014). Regarding exclusion criteria, we removed one participant with more than 40 % errors in the VAAST and nine other participants with exceptionally short response times (i.e., $\geq$ 30 % responses faster than 150 ms, the median value) in the reverse correlation (RC) task, leaving us with a sample of 138 participants ($M_{age} = 37.88$ years, $SD_{age} = 12.51$; 66 men and 72 women; similar to Rougier et al., 2021). This experiment had a 2 (movement: approach vs. avoidance) x 2 (framing: affiliative vs. aggressive approach) x 2 (background color: blue vs. yellow) x 2 (base-face group: Group 1 vs. Group 2) mixed design with movement varying within participants and the other factors between them. The last two factors served as controls for the analyses.

Procedure. We built the experiment with Psytoolkit (www.psytoolkit.org, Stoet, 2010, 2017) and administered it online. At the outset of the experiment, participants were told that the study aimed to investigate how people categorize others. They provided their consent before starting the experiment.

Step 1. Approach/Avoidance Training. At this stage, we randomly assigned participants to the affiliative approach or to the aggressive approach condition. Following Krieglmeyer and Deutsch (2013), we manipulated the framing associated with approach actions by means of instructions. Half of participants read that "approach means affiliation: It represents situations in which we approach for a positive verbal or physical interaction with the person in front of us" (affiliative approach condition). The other half read that "approach means to aggress: It represents situations in which we approach to verbally or physically aggress/attack the person in front of us" (aggressive approach condition). Because our focus was on manipulating the framing underlying approach (see Krieglmeyer & Deutsch, 2013), we did not manipulate distinct avoidance framings. Specifically, participants systematically read that "avoidance means to run away: It represents situations in which we run away/avoid the person in front of us". After receiving the instructions, participants performed the AAT.

We adapted the VAAST procedure to use it as an AAT, as in Rougier et al. (2021), see also Aubé et al., 2019). This task emulates the visual aspects associated with movements of the self in a virtual environment (i.e., a regular virtual street view giving a depth impression, see Fig. 1). Depending on the background color (blue vs. yellow) of the face appearing in the environment, participants had to perform approach or avoidance actions. Half of the participants had to approach blue-background faces and to avoid yellow background faces whereas the other half had the reverse instructions.

At the beginning of each trial, a white circle appeared at the center of the screen, informing participants that they could start the trial. As soon as participants pressed the start button (the H key on an AZERTY-keyboard), a fixation cross replaced the circle (for a random duration between 800 and 2000 ms), followed by a face (see Fig. 1). At the onset of the face, participants had to categorize it as a function of the background color as quickly and as accurately as possible by pressing the Y key (approach) or the N key (avoidance) of their keyboard. Depending on participants' approach vs. avoidance action, the visual environment changed (i.e., the background image and the target face). Specifically, the face was zoomed in vs. out (by approximately 13 %) and the visual background was replaced by a close-up vs. distant shot of its initial image. These visual manipulations gave participants the visual impression that they were moving toward or away from the face in the street environment. In case of an inaccurate response, a red cross ("X") appeared on the screen during 500 ms before the next trial.

In total, participants performed 192 trials over 16 faces (eight faces of Group 1 and 8 faces of Group 2) presented 12 times in a random order. Half of the faces (of Group 1 vs. Group 2 according to the condition) were approached and half of them were avoided. All faces were white male faces from the CaNAFF face database (Courset et al., 2018) used in Rougier et al. (2021). The two groups of faces did not differ significantly in terms of emotional neutrality, $t(14) = 1.26$, $p = .23$, $d = 0.63$, 95 % CI [-0.47; 1.73], approach/avoidance tendencies, $t(14) = 0.46$, $p = .65$, $d = 0.23$, 95 % CI [-0.85; 1.30], and attractiveness, $t(14) = 0.001$, $p = .99$, $d = 0.29$, 95 % CI [-1.07; 1.07]. To further minimize the idiosyncratic features of the faces, we slightly blurred the faces (Gaussian blur of 3.5 radius). We randomly exposed participants to a framing (affiliative vs. aggressive), a background color (blue vs. yellow) associated with the group (Group 1 vs. Group 2), and an instruction relative to the color (i.e., approach blue-background faces vs. avoid blue-background faces).

Step 2. Reverse Correlation. After the AAT, participants underwent a reverse correlation procedure adapted from Dotsch et al. (2008), see also Dotsch & Todorov, 2012; Dotsch et al., 2013), as in Rougier et al. (2021, Exp. 1). Before the task, participants were told that each group of faces presented previously (i.e., with a yellow vs. blue background) were physically very different from the other group and that, within each group, the faces also shared a series of physical characteristics, making them similar to each other. The alleged goal of the task was therefore to identify faces from the yellow- and blue-background groups. In one block, the face selection was about the yellow-background group and, in the other block, about the blue-background group. The block order was randomized. For each trial, two noisy faces appeared side-by-side on the screen and participants had the following instruction: "select the face that you think is the most similar to the blue [yellow-] - background group of faces [...] (i.e., the faces that you ran away from/avoided [affiliated with/approached; aggressed/attacked])". Participants selected the face by using the S and the L keys.

Each block comprised 200 trials and, accordingly, we generated a total of 200 pairs of noisy faces (100 original and 100 negative) using the R package *rcicr* version 0.3.4.1 (Dotsch, 2015) with the default settings. Noisy faces (512 × 512 grayscale pixels) consisted in a base image with superimposed random noise. We used two separate base images that were the average faces (i.e., the morph) of the two groups presented in the VAAST: One base image (Base Image 1) was the average face from Group 1 and the other base image (Base Image 2) was the average face from Group 2 (see Fig. 2). We then added a grey background to each base image, converted them to a grayscale and slightly blurred them (radius 5 pixels and Laplacian standard deviation of 5 pixels). Within a block, the base image always corresponded to the target group (e.g., when participants had to select the blue-background face, the base image was a morph of the blue-background faces seen in the VAAST).

Regarding the noise applied on the base image, we generated a different noise pattern for each trial, the set of noise patterns being the same for the two blocks and for all participants. For each pair, one stimulus consisted in the base image along with the original random noise, and the other was the base image with the negative (opposite) pattern of noise (see Fig. 2). We always presented the two images as pairs and the image with the original noise appeared randomly either on the right or the left side of the screen. We showed the pairs in a random order within each block. At the end of the experiment, participants answered demographic questions (age and gender) and were debriefed.

2.2.2. Part 2A: classification images ratings (condition-level) by independent judges

In Part 2A of the experiment, we tested whether the condition-level classification images obtained in Part 1 elicited different judgments. Our main hypothesis was a framing moderation on the AAT effect. Accordingly, in the affiliative condition, we expected the approach classification image to be evaluated more positively than the avoided one. We predicted that this difference would decrease or reverse in the aggressive condition.

Construction of the Classification Images. We constructed the classification images as a function of movement (i.e., approach vs. avoidance) and condition (affiliative vs. aggressive approach) for each base image (Base Image 1 vs. 2).[4] Specifically, we averaged all the selected noises for all the participants within each condition and we then superimposed this average noise to the corresponding base image, resulting in 8 condition-level classification images (see Fig. 3; scaling method = autoscale). To test whether background color (blue vs. yellow) and block order (first vs. second) affected our results, we also created classification images for each of these conditions (in addition to the main conditions of interest), resulting in 32 classification images.

Participants. One hundred Prolific Academic users ($M_{age} = 34.03$ years, $SD_{age} = 12.16$; 46 men, 52 women, and 2 responding 'other' – i.e., self-categorizing as neither a man nor a woman) took part in exchange for a monetary compensation (£5.00/h; same pre-screening criteria as in Part 1).

Procedure. We programmed the online study via Qualtrics. Before signing the informed consent, participants learned that the study was about face perception and that their task would be to evaluate 40 faces on a series of traits (i.e., the 8 condition-level classification images and the 32 classification images testing for the control variables). We used

[4] Data from different base images cannot be averaged together.

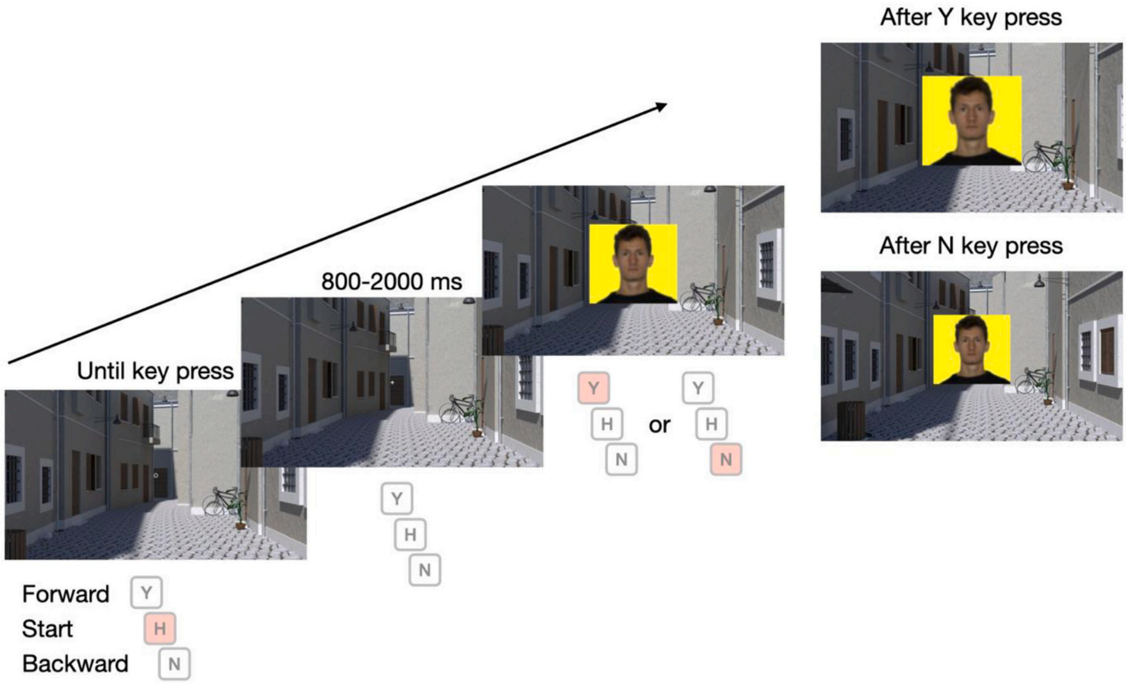

Fig. 1. Time Course of a Trial in the VAAST (from Rougier et al., 2021).
Note. The street image and the background area of the face were originally colored (in this example, the target face is presented with a yellow background). After pressing the H key (start) and upon appearance of the target face, the participant had to press the Y (approach) or the N key (avoidance) as a function of the face's background color.

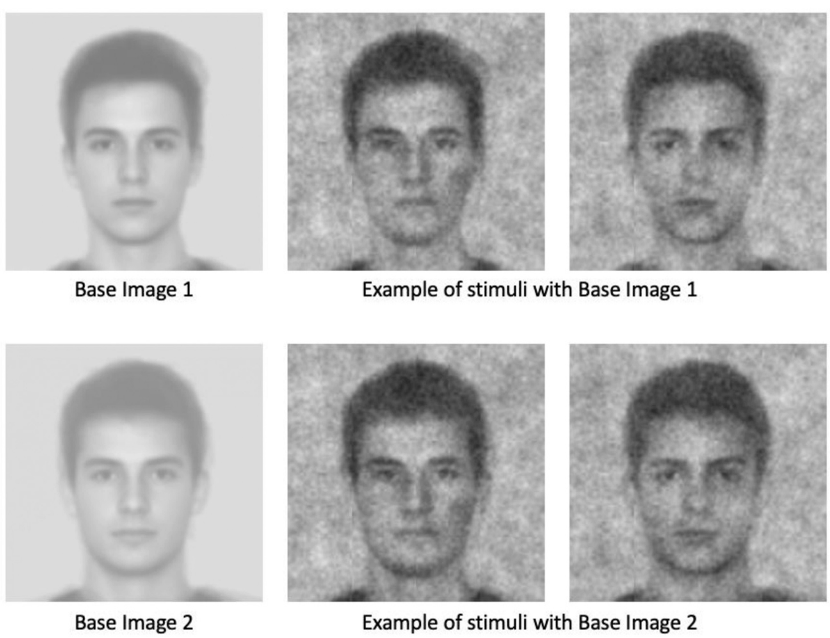

Fig. 2. Base Images and Associated Examples of Stimuli for a Given Noise in the Reverse Correlation Task.
Note. Left panel depicts the Base Image (1 on the top left and 2 bottom left). The two right panels depict pairs of images with opposite patterns of noise (from Rougier et al., 2021).

the traits *aggressiveness*, *trustworthiness*, and *criminality* used in Rougier et al. (2021), that is, traits typically related to the "approachability/avoidability" of a person (Krieglmeyer & Deutsch, 2013; Slepian et al., 2012; Wentura et al., 2000).

We presented the faces one by one in a random order and the aggressive, trustworthy, and criminal scales (on a continuous scale from 0 = *not at all* to 100 = *very much*) adjacent to each other in that order. We encouraged participants to answer as honestly and as spontaneously as possible. Finally, participants completed the same demographics as in Part 1.

2.3. Results

Because judges were the unit of analysis and rated all the classifi-

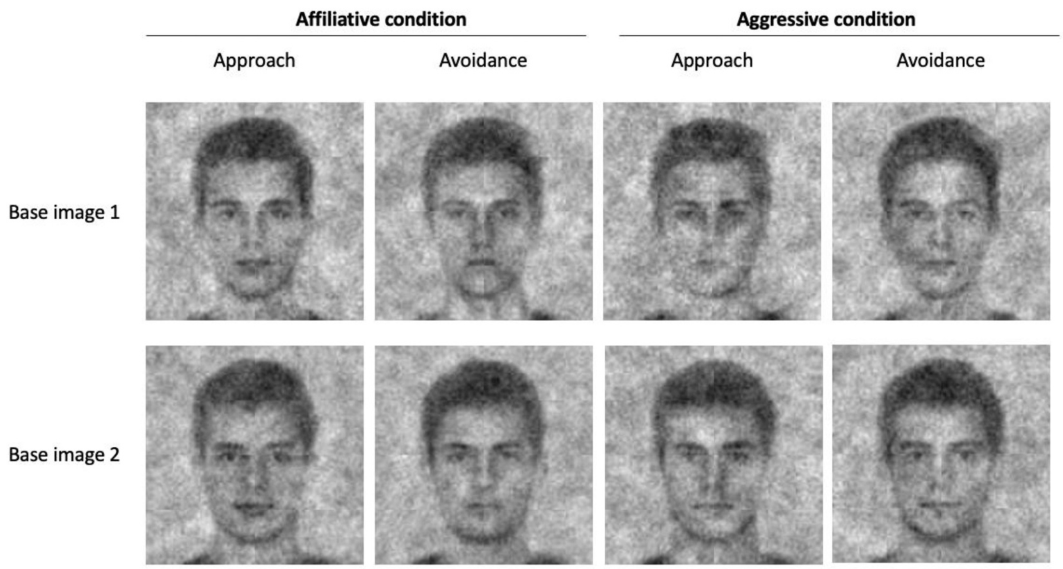

Fig. 3. Condition-Level Classification Images as a Function of the Condition (Affiliative vs. Aggressive Approach), the Movement (Approach vs. Avoidance), and the Base Image (Base Image 1 vs. 2).

cation images, we treated the movement and framing variables as within-judge factors in our analyses. We report both frequentist and Bayesian analyses. Indeed, Bayesian analyses allow to gauge the amount of evidence in favor of a movement by framing interaction (H_1) versus a systematic AAT effect (i.e., the absence of a movement by framing interaction; H_0) predicted by the motivational-systems and the self-anchoring accounts. We report the BF_{10} (evidence in favor of H_1) when the frequentist analysis reached significance (i.e., $p < .05$), and the BF_{01} (evidence in favor of H_0) when it did not. We carried out the analyses with the JZS default Bayes factor (*ttestBF* function) of the *BayesFactor* R package (version 0.9.12-4.2.; Morey & Rouder, 2015). Accordingly, the model used a "default" Cauchy prior distribution with the *r* scale $\sqrt{2}/2$ (Rouder et al., 2009). The interpretations of the *BF* are based on Lee and Wagenmakers's (2013) classification.

Overall, aggressiveness and criminality scores were highly correlated, $r = .74$, 95 % CI [.70; .77], $t(798) = 31.16$, $p < .001$, while trustworthiness was correlated more modestly with aggressiveness, $r = -.55$, 95 % CI [-.60; -.50], $t(798) = 18.65$, $p < .001$, and criminality, $r = -.59$, 95 % CI [-.63; -.55], $t(798) = 20.98$, $p < .001$. In line with Rougier et al. (2021), we computed a "positivity score" (α = .84), indicating the extent to which the judges perceived a face as being, on average, more trustworthy, less aggressive, and less criminal. Given that none of the control factors influenced the movement by framing interaction ($t(98) = 0.81$, $p = .42$, $dz = 0.08$, 95 % CI [-0.32; 0.48],[5] $BF_{01} = 6.53$, for the background color, $t(99) = 1.17$, $p = .24$, $dz = 0.12$, 95 % CI [-0.28; 0.51], $BF_{01} = 4.67$, for the block order), we excluded them from the analyses.

Our main analysis revealed the predicted movement (approach vs. avoidance) by framing (affiliative vs. aggressive) interaction, $t(99) = 16.15$, $p < .001$, $dz = 1.61$, 95 % CI [1.16; 2.07], $BF_{10} > 100$, (see Fig. 4, left panel). This interaction indicates that the AAT effect observed in the affiliative approach condition ($M_{App} = 70.24$, $SE = 1.33$; $M_{Av} = 40.36$, $SE = 1.34$) reversed in the aggressive approach condition ($M_{App} = 46.09$, $SE = 1.31$; $M_{Av} = 57.52$, $SE = 1.24$). Analysis of the simple effects showed that the movement effect was significant in the affiliative approach condition, $t(99) = 15.46$, $p < .001$, $dz = 1.55$, 95 % CI [1.09; 2.00], $BF_{10} > 100$, as well as in the aggressive approach condition, $t(99) = 10.03$, $p < .001$, $dz = 1.00$, 95 % CI [0.58; 1.42], $BF_{10} > 100$.

Moreover, the classification images associated with affiliative ($M = 70.24$, $SE = 1.33$) and aggressive approach ($M = 46.09$, $SE = 1.31$) differed from each other, $t(99) = 15.31$, $p < .001$, $dz = 1.53$, 95 % CI [1.08; 1.98], $BF_{10} > 100$. This was also the case for avoidance classification images in the affiliative ($M = 40.36$, $SE = 1.34$) and aggressive approach conditions ($M = 57.52$, $SE = 1.24$), $t(99) = 12.13$, $p < .001$, $dz = 1.21$, 95 % CI [0.78; 1.64], $BF_{10} > 100$.

2.3.1. Part 2B: classification images ratings (subgroup-level) by independent judges

In Part 2A, we observed the expected framing moderation when relying on condition-level classification images. However, the conclusions we can draw from these results remain limited. Indeed, recent work suggests that relying on condition-level classification images is problematic because condition-level images ignore the variability between face producers of a given condition (Cone et al., 2021). Ignoring this source of variability may increase Type I error rate (i.e., claiming that there is a significant difference between the classification images when there is in fact none).

One solution that has been recommended (Cone et al., 2021) and successfully implemented in the past (Rougier et al., 2021) is to rely on subgroup-level classification images, that is, images composed of randomly selected individual responses within a condition (e.g., for approach in the affiliative condition when using base image 1). Subgroup-level classification images would retain a substantial portion of individual-level variability (i.e., variability stemming from face producers) while allowing for a good signal-to-noise ratio, that is, ensuring a good "readability" of the classification images by the judges. One word of caution though, although subgroup-level classification images is in principle less prone to type I errors than condition-level images, further work still needs to address the extent to which the false positivity rate is reduced. In Part 2B, we adopted this strategy and tested whether the observed framing moderation replicates when relying on subgroup-level classification images.[6] Part 2B was preregistered, which included a priori theoretical reasoning, hypotheses, power estimation, procedure, and statistical analyses.

Construction of the Classification Images. We built each subgroup-level classification image using five individual-level data

[5] For this analysis, we removed one outlier with a score above 4 on studentized deleted residuals (McClelland, 2014).

[6] We thank an anonymous reviewer for the suggestion.

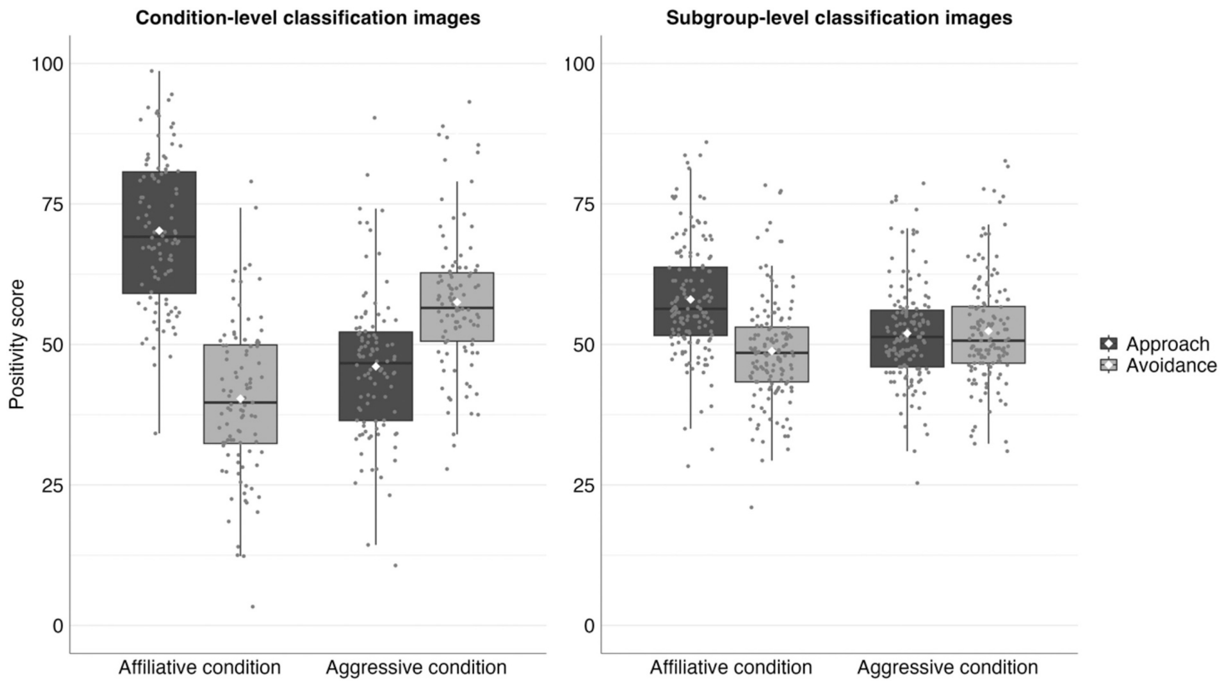

Fig. 4. Boxplot Representing the Positivity Score as a Function of Movement (Approach vs. Avoidance) and Condition (Affiliative Framing vs. Aggressive Framing) for condition-level classification images (left panel) and subgroup-level classification images (right panel).
Note. The positivity score represents the average score of trustworthiness, aggressiveness (reversed), and criminality (reversed). Lines inside the boxes represent the median values and white diamonds represent the mean values. The lower edges of the boxes represent the lower quartiles and the upper edges the upper quartile.

within the same experimental condition from the 2 (framing of approach: affiliative, aggressive) x 2 (movement: approach, avoidance) x 2 (base image: Base Image 1, Base Image 2). For each crossing of the three variables (8 conditions), we had a minimum of 32 face creators. Therefore, we randomly selected the data of 5 face creators (among the 32) for each crossing. As face creators went through two blocks in the reverse correlation task to select faces (one block for the approached group of faces and one block for the avoided group of faces), data of the 5 same participants were selected to compute the subgroup classification images of approach vs. avoidance within a condition.[7] We then averaged the classification images of these five participants to generate a subgroup classification image for approach and for avoidance (as we did previously for the condition-level classification images). This procedure was repeated until we generated 256 subgroup-level classification images, with half in the affiliative approach condition and half in the aggressive approach condition and within each, half of classification images were of approach and half of avoidance.[8]

Participants. One hundred and fifty Prolific Academic users (M_{age} = 39.22 years, SD_{age} = 12.94; 84 men, 63 women, and 3 responding 'other') took part in exchange for a monetary compensation (£7.00/h). Pre-screening criteria were the same as before except that the approval rate was set at 98 %.

Procedure. We programmed the online study on JsPsych (de Leeuw, 2015). The procedure was the same as in Part 2A except that each participant evaluated 40 subgroup-level classification images, with 5 images for each combination of framing, movement, and base image (random selection).

2.4. Results

Again, we considered judges as the unit of analysis and treated movement and framing variables as within-judge factors in our analyses. Based on our preregistered exclusion criterion, we excluded six participants showing less than 5 % of variance in their responses. Aggressiveness and criminality scores were highly correlated, r = .71, 95 % CI [.69; .72], t(5758) = 75.83, p < .001, while trustworthiness was correlated more modestly with aggressiveness, r = -.51, 95 % CI [-.53; -.49], t(5758) = 44.65, p < .001, and criminality, r = -.50, 95 % CI [-.52; -.48], t(5758) = 43.58, p < .001. As before, we computed a "positivity score" (α = .80).

The expected interaction between movement (approach vs. avoidance) and framing (affiliative vs. aggressive) emerged, t(142) = 9.57, p < .001, dz = 0.80, 95 % CI [0.61; 0.99], BF_{10} > 100 (see Fig. 4, right panel). This interaction indicates that the AAT effect was larger in the affiliative condition (M_{App} = 58.01, SE = 0.85; M_{Av} = 48.83, SE = 0.78) than in the aggressive one (M_{App} = 51.92, SE = 0.74; M_{Av} = 52.39, SE = 0.80). Whereas the simple effect of movement in the affiliative condition was significant, t(142) = 11.91, p < .001, dz = 1.00, 95 % CI [0.80; 1.20], BF_{10} > 100, it did not emerge in the aggressive condition, t(143) = 0.73, p = .46, dz = 0.06, 95 % CI [-0.10; 0.23], BF_{01} = 8.29. Crucially, the simple effect of framing was significant in the approach condition, t(143) = 9.21, p < .001, dz = 0.77, 95 % CI [0.58; 0.96], BF_{10} > 100. We also observed this effect in the avoidance condition, t(142) = 6.09, p < .001, dz = 0.51, 95 % CI [0.34; 0.69], BF_{10} > 100.

2.5. Discussion

Consistent with our hypothesis, the framing of approach moderated the AAT effect. This effect emerged for both condition-level and subgroup-level classification images. Moreover, Bayes factors indicated extreme evidence in favor of H_1 (framing-dependency of AAT effect). When we framed approach as affiliative, it led to a more positive visual

[7] This means that the approach and avoidance subgroup-level classification images within a condition (e.g., affiliative approach) came from the same sample of 5 participants. Of note, for a given subgroup image, when the Base Image 1 for approach (avoidance), it was the Base Image 2 for avoidance (approach).

[8] Within each condition, there was 32 images of approach with Base Image 1, 32 of approach with Base Image 2, 32 of avoidance with Base Image 1, and 32 of avoidance with Base Image 2.

representation than avoidance, replicating the standard AAT effect (Rougier et al., 2021, Exp. 1). In contrast, when we framed approach as aggressive, the observed effect tended to reverse in that the difference between approach and avoidance images only reached significance for condition-level classification images. Crucially, the aggressive approach led to a more negative facial representation of the approached group than the affiliative approach. Overall, this pattern corroborates our reasoning that the action framing moderates the outcome of AAT.

3. Registered Experiment 1

Our Preliminary Experiment constitutes initial suggestive evidence for the framing-dependency of AAT effects. To find convergent evidence, however, we aim to replicate these findings while addressing several limitations. First, although we did not manipulate the framing associated with avoidance, the classification image of the avoided group was more negative in the affiliative than in the aggressive approach condition. This implies that the effect of approach/avoidance behaviors on visual representations depends on the meaning conveyed by the other movement. To control for such source of variation, we induced only one movement per condition along with the same control action and compared them: affiliative approach, aggressive approach, and (generic) avoidance.

Second, as predicted by a grounded cognition perspective, aggressive approach and (generic) avoidance should both lead to negative visual representations but should manifest on different facial features. This should be the case because their respective multimodal representations include different visual characteristics of the distinct eliciting stimuli that triggered these actions. Ultimately, the biases on the visual representations should reflect different judgments on personality traits (e.g., lower trustworthiness ratings).[9] However, we only assessed the AAT effect on the visual representations by using sets of positive/negative traits. This did not allow us to test for differences between the visual representations resulting from aggressive approach and avoidance beyond a *mere evaluative* bias.

To test the specificity of both actions on facial features, we need to use traits that clearly relate to aggressive approach vs. avoidance (see Rougier et al., 2021, for a similar reasoning). To objectively assess which traits are usually most likely to trigger aggressive approach vs. avoidance, we conducted a pilot study where we asked participants to indicate whether a series of traits would likely lead to aggressive approach vs. to avoidance reactions (see Supplementary Materials for more information). We expected the visual representations of aggressive approach and avoidance to be different on these two categories of traits.

3.1. Power analysis and sample size

As data analyses are only performed on the judges' ratings, the mere information available for the face producers was the sample size per condition from previous similar experiments (i.e., $N = 71$ per condition in our Preliminary Experiment and $N = 93$ per condition in Rougier et al., 2021, Exp. 1). Relying on the largest sample size as a backup strategy, we aimed at recruiting a sample of $N = 300$ face producers (i.e., slightly above $N = 279 = 93 \times 3$ conditions) to safeguard from participant exclusion. Regarding our sample of judges, we relied on the effect size of the framing moderation obtained for the subgroup-level classification images (Part 2B; $dz = 0.80$) in our Preliminary Experiment. A similar sample as the one we used in the Preliminary Experiment (Part 2B; $N = 150$) would provide us with a power of 99.99 % to perform frequentist analyses (5 % false-positive rate in a two-tailed *t*-test with paired samples). Considering that we might end up with a smaller effect (because of procedural differences between current and past experiments such as the between-participants manipulation of approach and avoidance), we also report power analyses relying on the average effect size in social psychology, $dz = 0.42$ (Richard et al., 2003). A sample of $N = 150$ afforded 99.97 % power to detect this effect (5 % false-positive rate in a two-tailed *t*-test with paired samples).

As in the Preliminary Experiment, we aimed at performing Bayesian analyses to better gauge the evidence in favor of our hypotheses (i.e., the framing-dependency of the AAT effect and the qualitative differences between aggressively approached and avoided groups; H_1; BF_{10}) or the alternative effects (H_0; BF_{01}). To maximize our chances to reach strong evidence, we relied on a "sequential Bayes factor with maximal n" procedure (Schönbrodt & Wagenmakers, 2018). This procedure consists in keeping collecting data until one reaches substantial evidence in favor of H_1 or H_0 and/or until one achieves the maximal sample defined beforehand. We set the evidential thresholds as $BF_{10} = 30$ and $BF_{01} = 1/6$,[10] $n_{min} = 150$, $n_{max} = 230$. It means that, after having collected $n_{min} = 150$ – enough data for the frequentist analyses – we computed BF_{10} and BF_{01} after every 20 participants and stopped data collection when one of the three criteria was reached (i.e., $BF_{10} = 30$, $BF_{01} = 1/6$, $n_{max} = 230$).

To estimate the viability of our parameters, we performed simulations using the R package BFDA (Schönbrodt, 2016). For an expected effect size of $dz = 0.80$ under H_1 (lower bound of the expected effect), 100 % of all simulated studies hit the correct H_1 threshold (i.e., the true positive rate) and 0 % hit the wrong H_0 threshold (i.e, the false negative rate). We observed similar parameters when considering a much smaller effect size, such as $dz = 0.42$ (i.e., average effect size in social psychology; Richard et al., 2003) – 99.8 % of all simulated studies hit the correct H_1 threshold and 0 % hit the wrong H_0 threshold. Under H_0 ($dz = 0$), 87.4 % of all studies hit the correct H_0 threshold and 1.2 % hit the wrong H_1 threshold (i.e., the false positive rate). The remaining 12.6 % of studies stopped at n_{max} and remained inconclusive with respect to the a priori set thresholds.

3.2. Method

3.2.1. Part 1: creation of classification images

Participants and Design. Three hundred and one Prolific Academic users took part in this experiment ($M_{age} = 41.16$ years, $SD_{age} = 12.23$; 186 men, 108 women, and 7 responding 'other') in exchange for a monetary compensation (£ 7.00/h). Inclusion and exclusion criteria were similar to our Preliminary Experiment except that we set a more conservative threshold regarding the error rate in the VAAST (i.e., participants with an error rate of 30 % or more, $N = 14$) and the response time in the RC (i.e., participants with 30 % or more responses faster than 200 ms, $N = 24$; see Rougier et al., 2021). Additionally, we planned to exclude participants having close to zero variance (i.e., less than 5 % variation) in their responses in the reverse correlation task (i.e., participants almost always providing the same answer; see also Rougier et al., 2021) but there was none. This left us with a final sample of 263 participants. Finally, for each ANOVA, we excluded outliers having a score of at least 4 on studentized deleted residuals or a gap in terms of their *Cook's d* or leverage values (exclusions are documented in the analytic R scripts; Judd et al., 2011).

The experiment relied on a 3 (target action: affiliative approach vs. aggressive approach vs. avoidance) x 2 (background color: blue vs. yellow) x 2 (face-based group: Group 1 vs. Group 2) between participants design. The last two factors served as control in the analyses. For

[9] Although the mapping between facial features and personality traits is not 'one to one', it is well documented that individuals spontaneously infer personality traits based on people's faces and that these judgments are very quick and consensual (e.g., Willis & Todorov, 2006; Zebrowitz et al., 2003). For instance, specific sets of facial features (e.g., jaws thickness, femininity, babyfacedness) drive trustworthiness or dominance judgments in a consistent manner (Oosterhof & Todorov, 2008).

[10] Note that we used asymmetric boundaries to avoid false positive evidence (i.e., concluding to H_1 although H_0 is true).

the reverse correlation task, participants performed 250 trials but this time only in one block.

Procedure. We programed the experiment on jsPsych and administered it online (via the Prolific Academic platform).

Step 1. Approach/Avoidance Training. The AAT procedure was the same as in the Preliminary Experiment, except for the following series of changes. First, we manipulated the target action variable (i.e., affiliative approach, aggressive approach, or avoidance) between participants and contrasted it to a control action (a condition without distance variation). Half of all participants received the instruction to perform the target action (i.e., affiliative approach vs. aggressive approach vs. avoidance) toward blue-background faces and to perform the control action toward yellow-background faces; the other half received the reverse instructions. For the control action, participants had to press the start key to categorize the corresponding group of faces. When doing so, the visual environment remained static (i.e., no approach and avoidance visual feedback) and the face disappeared after a 650 ms delay following the key press.[11] Therefore, the control action involved the same time course as in approach or avoidance responses, the only difference being that no visual feedback of approach or avoidance was provided. Second, participants used the D (start), E (approach) and C (avoidance) keys of their keyboard.[12]

Step 2. Reverse Correlation. We used the same procedure as in the Preliminary Experiment except that: 1) there was no recall of which color group was associated with affiliative approach, aggressive approach, or avoidance to render our expectations (i.e., influence of the target action on the classification images) less salient and 2) we used a single base image (being the morphed face of all stimuli presented in the AAT) for practical reasons.[13] Participants thus either had to select the blue- or the yellow-background typical face – always corresponding to the target action (i.e., "Select the face that you think is the most similar to the group of faces that had a yellow [blue] background"). They did not have to select the typical face corresponding to the control action.

3.2.2. Part 2: classification images ratings by independent judges

To investigate whether the classification image of affiliative approach elicits more favorable evaluations than the classification image of aggressive approach, we used the same traits as in our Preliminary Experiment (i.e., trustworthiness, aggressiveness, and criminality). We also investigated whether the classification image of avoidance differed from the classification image of aggressive approach. We hypothesized that the former should show more bias on traits related to avoidance behaviors (i.e., "avoidance relevant" traits), whereas the latter should show more bias on traits related to aggressive approach (i.e., "aggressive approach relevant" traits).

Pilot Study. To obtain these two categories of traits, we conducted a Pilot Study (N = 52) in which participants rated to what extent they would react by avoiding vs. approaching aggressively a person holding a given target trait (see Supplementary Material for more information on the procedure and results). Participants also evaluated traits on their valence (how positive/negative is a trait) and "face relevance" (how easy it is to deduce a trait based on someone's face). The selected traits differed significantly with respect to avoidance (opportunistic, rough, dominant, racist) and aggressive approach (stingy, depressive, mediocre, weak), $t(6) = 3.47$, $p = .01$, $dz = 2.44$, 95 % CI [0.15; 4.72]. At the same time, they did not differ significantly on valence, $t(6) = 0.26$, $p = .80$, $dz = 0.18$, 95 % CI [-1.56; 1.92], and face relevance, $t(6) = 0.28$, $p = .79$, $dz = 0.20$, 95 % CI [-1.53; 1.93]. Interestingly, the obtained selection for traits relating to avoidance vs. aggressive approach is consistent with the idea that traits depicting general weakness provide the opportunity to aggress (i.e., because chances for a successful aggression are higher; see Muller et al., 2012). In contrast, traits that seem to trigger avoidance reactions rather denote a lack of opportunity – so that the target individuals should be avoided because of low chances of success to overcome the threat (e.g., Sell et al., 2014, 2008). Any difference with respect to these specific categories of traits would inform us on the selectivity of the effect that depends on the *type* of action beyond its mere negative valence. In line with a grounded cognition framework, this would suggest that different actions of the same valence influence traits of distinct categories.

Construction of the Classification Images. To visualize the facial representations associated with affiliative approach, aggressive approach, and avoidance, we averaged all the selected noises for all participants and each target action, and we then superimposed this average noise to the corresponding base image (constant scaling = .004).[14] This resulted in three condition-level classification images (see Fig. 5). Testing for the potential effect of control variables (here, background color and face-based group) requires computing classification images at each intersection of the control variables and the target action. We adopted this strategy in our Preliminary Experiment by relying on condition-level classification images (see also Rougier et al., 2021). In this and the following experiment, we relied on subgroup images instead, as it does not require generating additional images. Specifically, for each subgroup image, we randomly selected the data of five individuals from the same condition of target action and from the same condition of our control variables of background color and face-based group (constant scaling = .012). In total, we computed 300 subgroup-level classification images (random selection with replacement) with 100 images for each condition of target action, hence 25 for each combination of the control variables.

Cluster Test on Condition-Level Classification Images from Affiliative vs. Aggressive Approach Conditions. For Registered Experiments 1 and 2 we supplemented the evaluation of condition-level classification images by more formal, objective analyses (which were not registered).[15] Specifically, we performed cluster analyses that help identifying facial regions of interest on classification images that correlate with participants' response in the reverse correlation task (Chauvin et al., 2005).

Cluster analyses are based on the random field theory and enable to determine whether clusters of pixels exceed a (significance) threshold in a smooth Gaussian random field. We followed the procedure of Chauvin et al. (2005) using the 'stat4CI' toolbox on Matlab and adopted the same

[11] We initially registered a time of 300 ms but then realized that this duration was set at 650 ms in previous studies using the VAAST procedure (e.g., Rougier et al., 2018, Rougier et al., 2021). We thus decided to keep the 650 ms duration, as in this previous work, after receiving formal approval for this deviation from the Action Editor before data collection.

[12] We decided to change the response key to avoid any confound between approach/avoidance and "yes"/"no" answers (with Y = "yes" and N = "no").

[13] Using a single base image instead of two reduces the number of obtained classification images to one per target action – therefore, for the same number of face producers, classification images are of better quality.

[14] In this experiment and in Registered Experiment 2, we used a constant instead of the registered automatic scaling to compute the classification images. The constant refers to the pixel intensity of the noise when added to the base image: the lower the constant value, the higher the noise (but the less readable the face). When this parameter is left automatic, the resulting faces may appear very noisy. In our case, after having computed the classification images with automatic scaling, the faces looked very noisy and we suspected they would be hard to read for the judges. In line with Dotsch's (2015) recommendation, our strategy was thus to find the lowest constant that worked for all classification images within a certain level (i.e., condition- or subgroup-level).

[15] We thank the editor for this useful recommendation.

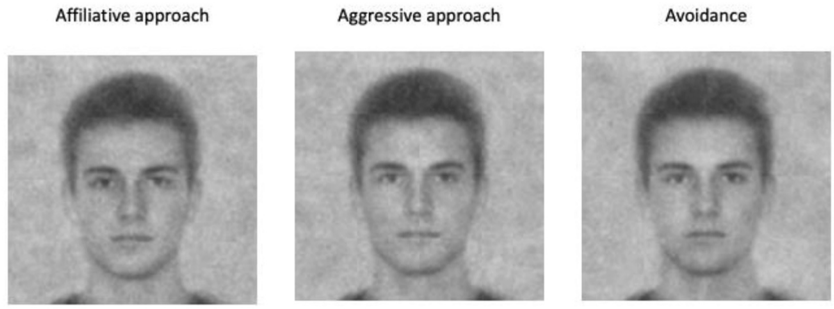

Fig. 5. Condition-Level Classification Images as a Function of the Target Action (Affiliative Approach vs. Aggressive Approach vs. Avoidance).

parameter values as in previous reverse correlation work (Dotsch & Todorov, 2012).[16] Green clusters indicate a positive correlation with face selection (i.e., significantly lighter pixels) whereas red clusters indicate negative correlation (i.e., significantly darker pixels).

As can be seen in Fig. 6, some face areas are statistically associated with participants' face selection that differ between approach conditions. Eye regions came across as statistically significant for both Affiliative and Aggressive approach classification images. The eye region (the area of the left eye to the nose and the eyebrow area) was significantly darker in the Affiliative approach condition (left image) whereas the eye area (inner area of the left eye) was lighter in the Aggressive approach condition (right image). This could indicate a greater sclera-pupil contrast in the Aggressive approach condition. Lighter areas on the lips also emerged as significant in the Aggressive approach condition, indicating more pursed-looking lips.

Participants. A total of 230 participants took part in the Part 2 of the experiment (M_{age} = 38.59 years, SD_{age} = 12.85; 105 men, 124 women, and 1 responding 'other'). As in Part 1, participants were recruited via Prolific Academic (same inclusion criteria). We excluded three participants with less than 5 % variation in their responses.

Procedure. We programed the experiment on JsPsych. We informed participants that the study was about face perception and that their task would be to evaluate faces on several dimensions. We divided the rating procedure into four blocks.

In a first block, participants evaluated the three condition-level classification images on the main dimensions of interest, that is, trustworthiness, aggressiveness, and criminality, to test for the framing moderation. The second block was identical except participants evaluated a subset of 24 subgroup-level classification images. The subgroup images were randomly selected from the pool of 300 subgroup images with the provision that each participant evaluated two images for each combination of movement, background color, and face-based group.

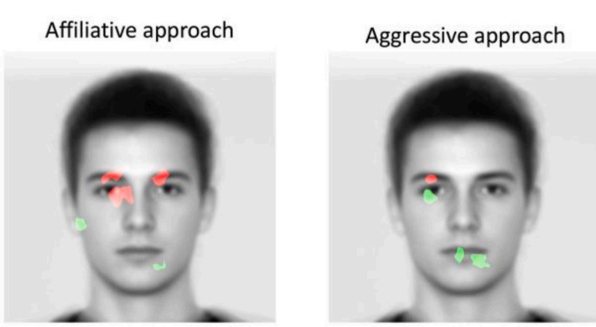

Fig. 6. Significant Clusters in Condition-Level Classification Images as a Function of the Approach (Affiliative Approach vs. Aggressive Approach).

Blocks 3 and 4 followed the same structure (first, rating of the condition-level and second, rating of the subgroup-level images) except participants evaluated the classification images on traits relating to avoidance (opportunistic, rough, dominant, racist) and to aggressive approach (stingy, depressive, mediocre, weak).[17] Subgroup-level classification images were identical between Block 2 and 4 for a given participant. The comparison of interest focused on classification images relating to aggressive approach and avoidance, but for exploratory purposes we also asked participants to rate the visual representation of affiliative approach.

Before the rating task within each block, we briefly displayed the three condition-level classification images (in Blocks 1 and 3) or a sample of subgroup-level classification images (in Blocks 2 and 4) for 2 s each, to help participants better gauge the similarities and differences between them. During the rating phase, we presented each face one by one in a random order within each block. The rating scales (Likert scale from 0 = *not at all* to 100 = *very much* with a 10-point gap between each response option) relative to each trait/emotion were presented one after the other on a vertical axis, always in the same order for a given participant (but the presentation order was randomized between participants).[18] Participants were prompted to answer as honestly and as spontaneously as possible. Finally, participants answered the same demographics as in Part 1.

3.3. Results

As in the Preliminary Experiment, our analyses relied on judges' ratings of condition- and subgroup-level classification images.

Testing the Framing Moderation. Overall, aggressiveness and criminality scores were highly correlated, r = .82, 95 % CI [.81; .82], $t(6127) = 110.62$, $p < .001$, while trustworthiness was correlated more modestly with aggressiveness, r = -.49, 95 % CI [-.51; -.47], $t(6127) =$ 44.12, $p < .001$, and criminality, r = -.49, 95 % CI [-.51; -.47], $t(6127) =$ 43.97, $p < .001$. We computed a "positivity score" as the average score of trustworthiness, aggressiveness (reversed), and criminality (reversed)

[16] This entailed (a) smoothing the classification noise pattern with a Gaussian filter (σ = 4 pixels), (b) selecting the face area with an oval-shaped mask (i.e., to only focus on the relevant area of the classification image), and (c) Z-transforming the classification image. Then, with two-tailed cluster tests, we identified which luminance variation in pixels positively (for light pixels) or negatively (for dark pixels) predicted face selection ($t > |2.3|$, $p < .05$) for each classification image.

[17] We adopted a block design because, whereas Blocks 1 and 2 are about the test of our main hypothesis (i.e., the framing moderation), the following blocks assess more secondary predictions (i.e., comparison of the type of facial bias on the visual representation resulting from aggressive approach vs. avoidance).

[18] We had to use a Likert scale (from 0 = *not at all* to 100 = *very much* with a 10-point gap between response options) instead of the registered continuous scale (from 0 to 100) because we found out that the implementation of a continuous scale in a multiple rating procedure is not supported by the programming software that we used (JsPsych).

ratings for each condition and subgroup classification image (α = .82).

We treated the target action variable as a within-judge factor. For each analysis, we indicate frequentist (*t*-tests) and Bayesian (BF_{10} or BF_{01} as a function of the *t*-tests significance) statistical indices as in the Preliminary Experiment (same priors and method). We performed all tests separately for both condition-level and subgroup-level classification images. When testing the effects for subgroup images, we also controlled for the potential effect of control variables (background color and face-based group).[19] We used as set of orthogonal contrast codes to test the framing-dependency of AAT effects.

The first contrast (C1 = 1/2 x affiliative approach - 1/2 x aggressive approach +0 x avoidance) compared the evaluation of the affiliative approach classification image(s) with the evaluation of the aggressive approach classification image(s). In line with a framing-dependency hypothesis, ratings of condition-level classification images significantly differed from each other, $t(225) = 3.20$, $p = .002$, $dz = 0.21$, 95 % CI [0.08; 0.35], $BF_{10} = 10.56$. The affiliative approach classification image was evaluated more positively than the aggressive approach classification image ($M_{affiliative} = 61.01$, $SE = 1.21$; $M_{aggressive} = 57.80$, $SE = 1.17$, see Fig. 7). This effect replicated for subgroup-level classification images ($M_{affiliative} = 57.89$, $SE = 0.44$; $M_{aggressive} = 57.19$, $SE = 0.46$), $t(225) = 2.05$, $p = .041$, $dz = 0.14$, 95 % CI [0.01; 0.27], $BF_{10} = 0.59$ (see Fig. 7).[20]

The second contrast (C2 = 1/3 x affiliative approach +1/3 x aggressive approach - 2/3 x avoidance) tested whether the evaluations of classification images in the affiliative and aggressive approach pooled together differed from the evaluation of the classification image in avoidance. Ratings of the condition-level classification images in affiliative and aggressive approach considered jointly ($M_{App} = 59.40$, $SE = 1.09$) were significantly higher than ratings of the classification image in avoidance ($M_{Av} = 52.91$, $SE = 1.19$), $t(225) = 7.77$, $p < .001$, $dz = 0.52$, 95 % CI [0.38; 0.66], $BF_{10} > 100$ (see Fig. 7). This effect replicated for subgroup-level classification images ($M_{App} = 57.54$, $SE = 0.81$; $M_{Av} = 55.37$, $SE = 0.88$), $t(223) = 4.77$, $p < .001$, $dz = 0.32$, 95 % CI [0.18; 0.45], $BF_{10} > 100$.[21] In other words, judges rated the approach classification images more positively than the classification images in avoidance (see Fig. 7).

Considering the observed pattern of results (see Fig. 7), we decided to conduct an additional, non-registered, analysis comparing aggressive approach and avoidance (contrast: 0 x affiliative approach +1/2 x aggressive approach – 1/2 x avoidance). Ratings of the condition-level classification image were significantly more positive for aggressive approach than for avoidance, $t(225) = 5.25$, $p < .001$, $dz = 0.35$, 95 % CI [0.21; 0.48], $BF_{10} > 100$. This effect replicated for subgroup-level classification images, $t(224) = 3.51$, $p < .001$, $dz = 0.23$, 95 % CI [0.10; 0.37], $BF_{10} = 27.68$.

Comparing Facial Features of Avoidance and Aggressive Approach. Finally, we tested whether the visual representations of avoidance and aggressive approach differed in terms of specific facial features, for both condition-level and subgroup-level classification images. To do so, we computed a negativity avoidance score (average score of opportunistic, rough, dominant, racist; α = .80) and a negativity aggressive approach score (average score of stingy, depressive, mediocre, weak, α = .78) for each classification image. Scores moderately correlated, $r = .62$, 95 % CI [.61; .64], $t(6127) = 62.47$, $p < .001$, and both the avoidance score, $r = -.38$, 95 % CI [-.40; -.36], $t(6127) = 32.00$, $p < .001$, and the aggressive approach score, $r = -.37$, 95 % CI [-.39; -.35], $t(6127) = 31.20$, $p < .001$, correlated negatively and moderately with the positivity score. Given that we compared only two types of visual representations, the movement was contrast-coded (aggressive approach: −0.5; avoidance: +0.5) as was the type of trait (aggressive approach relevant: −0.5; avoidance relevant: +0.5).

The interaction between the movement and the type of trait did not emerge, both when considering condition-level classification images, $t(225) = 1.05$, $p = .29$, $dz = 0.07$, 95 % CI [-0.06; 0.20], $BF_{01} = 7.80$, and subgroup-level classification images, $t(225) = 1.74$, $p = .084$, $dz = 0.12$, 95 % CI [-0.02; 0.25], $BF_{01} = 3.06$. On average, classification images received higher ratings on avoidance relevant traits than on aggressive approach relevant traits at the condition-level, $t(224) = 4.72$, $p < .001$, $dz = 0.32$, 95 % CI [0.18; 0.45], $BF_{10} > 100$, and subgroup-level, $t(224) = 3.58$, $p < .001$, $dz = 0.24$, 95 % CI [0.11; 0.37], $BF_{10} = 34.74$ (see Fig. 8). Results for the affiliative approach condition are reported as Supplementary Material (see Table S1).

3.4. Discussion

In Registered Experiment 1, we observed that classification images in the affiliative approach condition came across more positively than the ones in the aggressive approach condition. However, it should be noted that the Bayes factor indicated strong evidence ($BF_{10} > 10$) for condition-level classification images, and only anecdotal evidence for the subgroup-level images ($BF_{10} < 3$). Interestingly, classification images in the affiliative and aggressive approach considered jointly received more positive ratings than classification images in the avoidance condition. Taken together, these findings are consistent with the framing-dependency of AAT effects by controlling for potential contrast effects between target actions.

A differential effect of classification images of aggressive approach and avoidance on traits pertaining to aggressive approach vs. avoidance – that is, a trait-specificity effect – did not emerge. We will return to this finding in the General Discussion.

4. Registered Experiment 2

The goal of Experiment 2 was to provide a test of the unique role of experiencing the approach/avoidance training to contrast a grounded cognition account with the inferential account as well as to rule out explanations in terms of demand effects of the Preliminary Experiment and Registered Experiment 1.

Specifically, we compared a VAAST condition, in which participants performed the approach/avoidance training, to a mere instructions condition, in which they did not. The AAT condition was the same as in Experiment 1, except that we focused on the difference between affiliative approach and aggressive approach (both contrasted to a control action). In the mere instructions condition, participants received the same instructions as in the VAAST but they did not perform the actual training. Based on a grounded cognition account, we predicted a larger difference between the two visual representations resulting from affiliative and aggressive approach in the VAAST condition (i.e., when

[19] The two control variables were contrast-coded (−0.5; +0.5).

[20] The background color moderated C1, $t(223) = 3.27$, $p = .001$, $dz = 0.22$, 95 % CI [0.09; 0.35], $BF_{10} = 13.19$, so that the effect of C1 was significant and more visible on for the Blue background condition, $t(223) = 3.57$, $p < .001$, $dz = 0.24$, 95 % CI [0.11; 0.37], $BF_{10} = 33.64$ ($M_{affiliative} = 59.29$, $SE = 0.62$; $M_{aggressive} = 57.40$, $SE = 0.63$) than the Yellow background condition, $t(223) = 1.26$, $p = .21$, $dz = 0.08$, 95 % CI [-0.05; 0.22], $BF_{01} = 6.16$ ($M_{affiliative} = 56.49$, $SE = 0.61$; $M_{aggressive} = 56.98$, $SE = 0.61$). The face-based group significantly moderated C1, $t(225) = 3.04$, $p = .003$, $dz = 0.20$, 95 % CI [0.07; 0.33], $BF_{10} = 6.51$, so that the effect of C1 was significant and more visible on faces of Group 2, $t(225) = 3.72$, $p < .001$, $dz = 0.25$, 95 % CI [0.12; 0.38], $BF_{10} = 57.39$ ($M_{affiliative} = 59.49$, $SE = 0.61$; $M_{aggressive} = 57.21$, $SE = 0.61$) than faces of Group 1, $t(225) = 1.09$, $p = .28$, $dz = 0.07$, 95 % CI [-0.06; 0.20], $BF_{01} = 7.51$ ($M_{affiliative} = 56.29$, $SE = 0.62$; $M_{aggressive} = 57.17$, $SE = 0.63$).

[21] The background color moderated C2, $t(220) = 2.20$, $p = .03$, $dz = 0.15$, 95 % CI [0.02; 0.28], $BF_{10} = 0.80$, so that the effect of C1 was significant and more visible for the Blue background condition, $t(220) = 4.86$, $p < .001$, $dz = 0.33$, 95 % CI [0.19; 0.46], $BF_{10} > 100$ ($M_{App} = 58.35$, $SE = 0.44$; $M_{Av} = 55.46$, $SE = 0.64$) than the Yellow background condition, $t(220) = 1.49$, $p = .14$, $dz = 0.10$, 95 % CI [-0.03; 0.23], $BF_{01} = 4.45$ ($M_{App} = 56.74$, $SE = 0.43$; $M_{Av} = 55.27$, $SE = 0.65$). The face-based group did not significantly moderate C2, $t(222) = 0.73$, $p = .46$, $dz = 0.05$, 95 % CI [-0.08; 0.18], $BF_{01} = 10.25$.

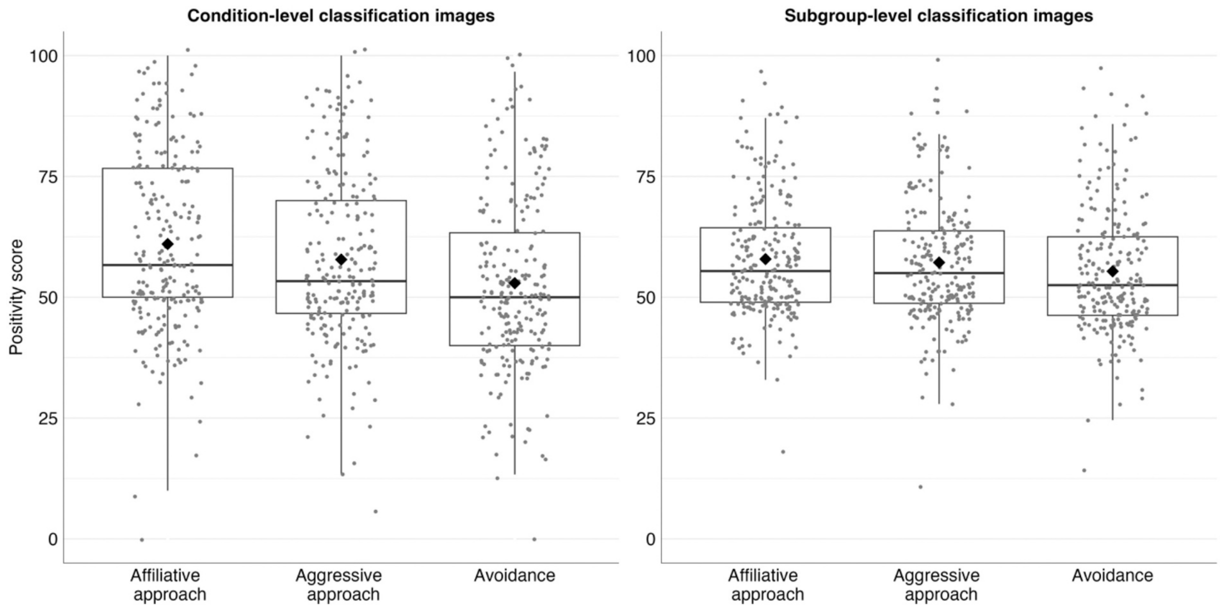

Fig. 7. Boxplot Representing the Positivity Score as a Function of the Target Action (Affiliative Approach vs. Aggressive Approach vs. Avoidance) for condition-level classification images (left panel) and subgroup-level classification images (right panel).
Note. The positivity score corresponds to the average score of trustworthiness, aggressiveness, and criminality, the last two scores being reversed. Lines inside the boxes represent the median values and diamonds represent the mean values. The lower edges of the boxes represent the lower quartiles and the upper edges the upper quartiles.

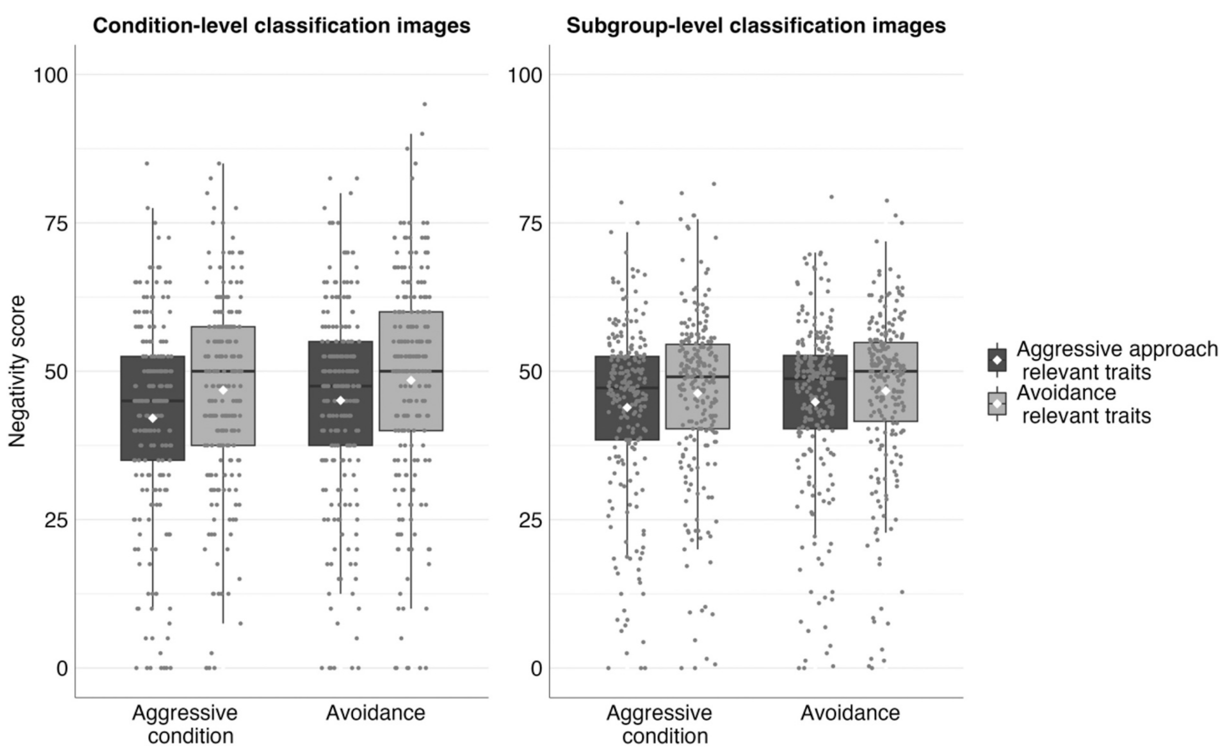

Fig. 8. Boxplot Representing Ratings as a Function of the Target Action (Aggressive Approach vs. Avoidance) and the Type of Trait (Avoidance Relevant vs. Aggressive Approach Relevant) for condition-level classification images (left panel) and subgroup-level classification images (right panel).
Note. Lines inside the boxes represent the median values and white diamonds represent the mean values. The lower edges of the boxes represent the lower quartiles and the upper edges the upper quartiles. In exploratory analyses reported as Supplementary Materials (see Table S1 and Fig. S1), we tested the differences between the visual representations of affiliative approach and aggressive approach, and between affiliative approach and avoidance – again, for both condition and subgroup classification images.

participants actually experience a training session) than in the mere instructions condition (i.e., when participants receive mere information about a prospective training session).

4.1. Power analysis and sample size

In line with Registered Experiment 1 (i.e., $N = 100$ face producers per condition), this experiment involved four between-participant conditions for Part 1 (see participants and design sections), so a sample size $N = 400$. This sample size for the producers exceeds the one used in the most similar experiment comparing the AAT and the mere instructions conditions ($N = 80$ per condition; Rougier et al., 2021, Exp. 3A).

Regarding the judges' sample size, as in Experiment 1, $N = 150$ judges allowed to secure 99.99 % power for the frequentist analyses (with $dz = 0.80$ and 5 % false-positive rate in a two-tailed t-test with paired samples). Again, we used a "sequential Bayes factor with maximal n" method to gather strong evidence in favor of either H_1 or H_0 (same parameters and procedure as in Experiment 1; Schönbrodt & Wagenmakers, 2018).

4.2. Method

4.2.1. Part 1: creation of classification images

Participants and Design. Four hundred students ($M_{age} = 21.43$ years, $SD_{age} = 2.47$; 126 men, 271 women, and 3 responding 'other') took part in this experiment in exchange for course credits or a monetary compensation (5€). We had no inclusion criteria and exclusion criteria were the same as in Registered Experiment 1, Part 1. Thus, we excluded participants with an error rate of 30 % or more in the VAAST ($N = 1$) and participants with 30 % or more responses faster than 200 ms ($N = 1$). No participant provided responses with less than 5 % of variation in the reverse correlation task but we removed participants who failed to remember correctly the approach/avoidance instructions ($N = 23$; see also Van Dessel et al., 2015). This left us with a final sample of 375 participants.

The design and procedure were the same as in Registered Experiment 1, except for the target action variable that only coded for affiliative vs. aggressive approach and the condition variable that coded for the VAAST vs. mere instructions manipulation. The experiment thus relied on a 2 (target action: affiliative approach vs. aggressive approach) x 2 (condition: VAAST vs. mere instructions) x 2 (background color: blue vs. yellow) x 2 (face-based group: Group 1 vs. Group 2) mixed design with the target action and condition variables varying between participants. The design of the VAAST was the same as in Experiment 1. To ensure that participants were equally familiar with the blue- and yellow-background faces in the mere instructions condition, participants underwent a control task (hereafter "categorization task") in lieu of the AAT (see Rougier et al., 2021; Exp. 3A). The design was the same as the AAT except that we used instructions relative to the key press (e.g., press the S key for blue-background faces) instead of approach/avoidance – participants did not perform any approach/avoidance action.

Procedure. We programed the experiment on jsPsych and it took place in a lab of the catholic University of Louvain. The procedure was comparable to Registered Experiment 1, except for a few changes related to the condition variable.

Step 1. Approach/Avoidance Induction. We randomly assigned participants to the VAAST or to the mere instructions condition. In the VAAST condition, the procedure was the same as in Registered Experiment 1 except that we did not include an avoidance condition.

In the mere instructions condition, participants started with a control categorization task. This task was the same as the VAAST (e.g., same visual environment) with the difference that participants categorized the blue- and yellow-background faces with the S and F keys and they used the D key as the start key. Moreover, there was no mention of approach/avoidance actions in the task and pressing the categorization keys was not associated with any visual feedback of approach/avoidance. As in Registered Experiment 1 for the control action, after pressing the S or F keys, the face remained static during 300 ms before disappearing. Because participants saw the two groups of faces the same number of times as in the VAAST and the trial time course was the same, the duration of the two tasks was highly similar. After the categorization task, participants received the same approach/avoidance instructions as in the VAAST condition. Specifically, we informed them that they would have to perform an approach/avoidance task (at the end of the study) and that their task would be to approach blue-background faces and to avoid yellow-background faces (or the reverse). Similar to the VAAST condition, we insisted on the importance of remembering these instructions. In fact, they never performed the AAT.

Step 2. Reverse Correlation. After the AAT, all participants underwent the same reverse correlation procedure as in Registered Experiment 1: They selected, within one block, the face that looked the most like the group of faces associated with approach (i.e., the target action).

Step 3. Instructions Check. After the reverse correlation task, participants reported the approach/avoidance instruction given to them (three response options: "Approach blue-background and remain static for yellow-background faces," "Approach yellow-background and remain static for blue-background faces," and "I do not remember"). Then, they completed the same demographics as before and were debriefed.

4.2.2. Part 2: classification images ratings by independent judges

In Part 2 of the experiment, we tested whether the predicted difference in evaluations between the two target actions (i.e., a more positive visual representation for the affiliative approach than the aggressive approach) was larger when participants truly experienced the AAT (i.e., the VAAST) as compared to when they received mere instructions. To test this hypothesis, we used the same set of traits as in Registered Experiment 1 (trustworthiness, aggressiveness, and criminality).

Construction of the classification images. To visualize the mental representations associated with the target action (affiliative approach vs. aggressive approach) as a function of the condition (VAAST vs. mere instructions), we computed the four condition-level classification images (constant scaling = .004; see Fig. 9). Regarding subgroup-level classification images, we randomly selected the data of five individuals within a given condition of target action and control variables of background color and face-based group (constant scaling = .012). In total, we computed 400 subgroup-level classification images (random selection with replacement) with 100 images for each condition of target action, hence 25 for each crossing of the control variables.

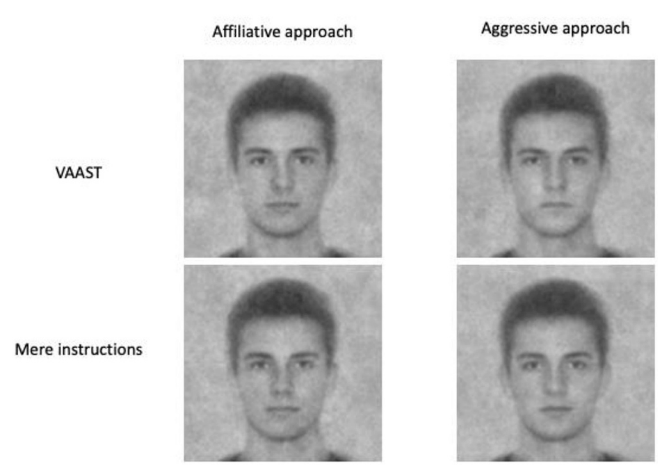

Fig. 9. *Condition-Level Classification Images as a Function of the Target Action (Affiliative Approach* vs. *Aggressive Approach) and the Condition (VAAST* vs. *Mere Instructions).*

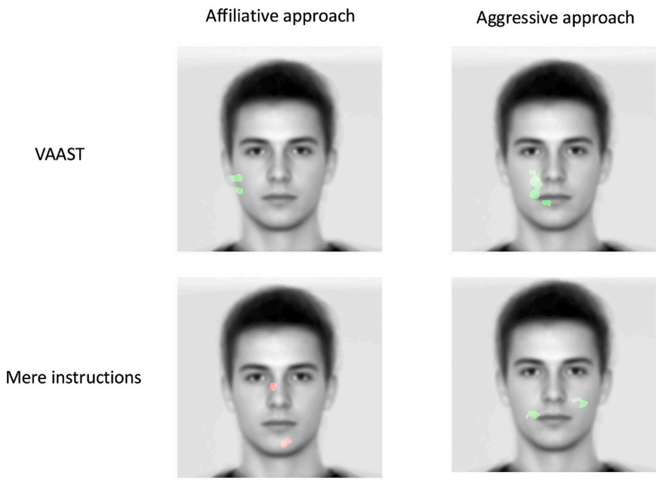

Fig. 10. Significant Clusters in Condition-Level Classification Images as a Function of the Approach (Affiliative Approach vs. Aggressive Approach) and the Condition (VAAST vs. Mere Instructions).

Cluster Test on Condition-Level Classification Images from Affiliative vs. Aggressive Approach Conditions. We performed the same (non-pre-registered) cluster analyses as in Registered Experiment 1. As can be seen in Fig. 10, some face areas were statistically associated with participants' face selection that seemed to differ between conditions. The mouth and cheek regions came across as significantly lighter for the Aggressive approach condition (see Fig. 10, top and bottom right images). This could indicate faces with more pursed-looking lips. Regarding the affiliative approach, the left (more external) cheek area was also significantly lighter in the VAAST condition (Fig. 10, top left image) and an area next to the upper zone of the nose and on the chin was darker in the mere instruction condition (bottom left image).

Participants. A total of 150 Prolific Academic users took part (M_{age} = 35.83 years, SD_{age} = 11.97; 70 men, 78 women, and 2 responding 'other').[22] Inclusion and exclusion criteria were the same as in Registered Experiment 1. We excluded three participants with less than 5 % variation in their responses.

Procedure. Before signing the informed consent, participants learned that the study was about face perception and that their task would be to evaluate faces on trustworthiness, aggressiveness, and criminality. The procedure comprised two blocks. In a first block, participants evaluated the four condition-level classification images and in a second block, they evaluated a sample of 32 subgroup-level images. The subgroup images were randomly selected in the pool of 400 subgroup images in such a way that each participant evaluated two images for each crossing of target action, condition, background color, and face-based group.

Before each block, we briefly displayed classification images as in Registered Experiment 1. We presented the faces one by one and the scales (Likert-type scales from 0 = *not at all* to 100 = *very much* with a 10-point gap between response options) relative to each trait were presented adjacent to each other on the vertical axis, always in the same order for a given participant. Participants were encouraged to answer as honestly and as spontaneously as possible. Finally, participants answered the same demographics as in Part 1.

4.3. Results

As previously, our analyses relied on judges' ratings of condition- and subgroup-level classification images. Overall, aggressiveness and criminality scores were highly correlated, r = .77, 95 % CI [.75; .78], $t(5290) = 86.68$, $p < .001$, while trustworthiness was correlated more modestly with aggressiveness, r = -.41, 95 % CI [-.43; -.39], $t(5290) = 32.78$, $p < .001$, and criminality, r = -.41, 95 % CI [-.43; -.38], $t(5290) = 32.34$, $p < .001$. We computed a "positivity score" as the average score of trustworthiness, aggressiveness (reversed), and criminality (reversed) ratings for the four classification images ($\alpha = .77$).

We treated the target action (aggressive approach: −0.5, affiliative approach: +0.5) and condition (mere instructions: −0.5, VAAST: +0.5) variables as within-judge factors. As in Experiment 1, we reported both frequentist (t-tests) and Bayesian (BF_{10} or BF_{01}) statistical indices for each analysis. We performed the analyses for both condition-level and subgroup-level classification images and we tested over subgroup images whether the main effects remained significant when controlling for the variables of background color and face-based group.

The interaction between the target action and the condition was significant for the condition-level classification images, $t(146) = 10.21$, $p < .001$, $dz = 0.84$, 95 % CI [0.65; 1.03], $BF_{10} > 100$ (see Fig. 11), indicating that the difference between classification images in the VAAST condition ($M_{affiliative} = 63.36$, $SE = 1.41$; $M_{aggressive} = 43.31$, $SE = 1.26$) was significantly larger than in the mere instructions condition ($M_{affiliative} = 59.30$, $SE = 1.27$; $M_{aggressive} = 59.93$, $SE = 1.30$). This pattern replicated for subgroup-level classification images (VAAST condition: $M_{affiliative} = 55.81$, $SE = 0.51$; $M_{aggressive} = 49.63$, $SE = 0.53$; Mere instruction condition: $M_{affiliative} = 56.76$, $SE = 0.51$; $M_{aggressive} = 55.72$, $SE = 0.51$), $t(146) = 5.62$, $p < .001$, $dz = 0.47$, 95 % CI [0.29; 0.64], $BF_{10} > 100$.[23]

In line with a framing-dependency hypothesis, analyses in the VAAST condition revealed that the condition-level classification image of affiliative approach received more positive ratings than the one of aggressive approach, $t(146) = 12.13$, $p < .001$, $dz = 1.00$, 95 % CI [0.80; 1.20], $BF_{10} > 100$. This effect also emerged for subgroup-level classification images, $t(146) = 8.24$, $p < .001$, $dz = 0.68$, 95 % CI [0.50; 0.86], $BF_{10} > 100$. In the mere instructions condition, classification images did not significantly differ at the condition-level, $t(146) = 0.50$, $p = .61$, $dz = 0.04$, 95 % CI [-0.12; 0.20], $BF_{01} = 9.61$. However, the effect emerged at the subgroup-level, $t(146) = 1.99$, $p = .048$, $dz = 0.16$, 95 % CI [0.00; 0.33], $BF_{10} = 0.63$.

4.4. Discussion

In Registered Experiment 2, we obtained a larger framing-dependency effect, that is a larger difference between affiliative vs. aggressive approach, in the VAAST condition than in the mere instructions condition. In the latter condition, we only observed a weak framing effect for classification images at the subgroup-level, but not the condition-level. These results are in line with the idea that the AAT effect is framing-dependent and that this effect hinges on the experiential aspect of the training.

[22] We chose to use an online sample of participants to accelerate the data collection procedure, given that there are no a priori reasons to believe that online independent judges would produce different results.

[23] The background color significantly moderated the observed interaction, $t(146) = 7.58$, $p < .001$, $dz = 0.63$, 95 % CI [0.45; 0.80], $BF_{10} > 100$, so that the interaction effect was larger for the blue background (VAAST: $M_{affiliative} = 58.57$, $SE = 0.70$; $M_{aggressive} = 47.77$, $SE = 0.76$; Mere instructions: $M_{affiliative} = 53.87$, $SE = 0.70$; $M_{aggressive} = 54.80$, $SE = 0.74$) than the yellow background (VAAST: $M_{affiliative} = 53.04$, $SE = 0.72$; $M_{aggressive} = 51.50$, $SE = 0.73$; Mere instructions: $M_{affiliative} = 59.65$, $SE = 0.71$; $M_{aggressive} = 56.65$, $SE = 0.70$). The interaction was significant for the blue background, $t(146) = 8.60$, $p < .001$, $dz = 0.71$, 95 % CI [0.53; 0.89], $BF_{10} > 100$, but not the yellow background, $t(146) = 1.26$, $p = .21$, $dz = 0.10$, 95 % CI [-0.06; 0.27], $BF_{01} = 5.02$. The face-based group did not significantly moderate the interaction, $t(146) = 1.90$, $p = .06$, $dz = 0.16$, 95 % CI [-0.01; 0.32], $BF_{01} = 0.71$.

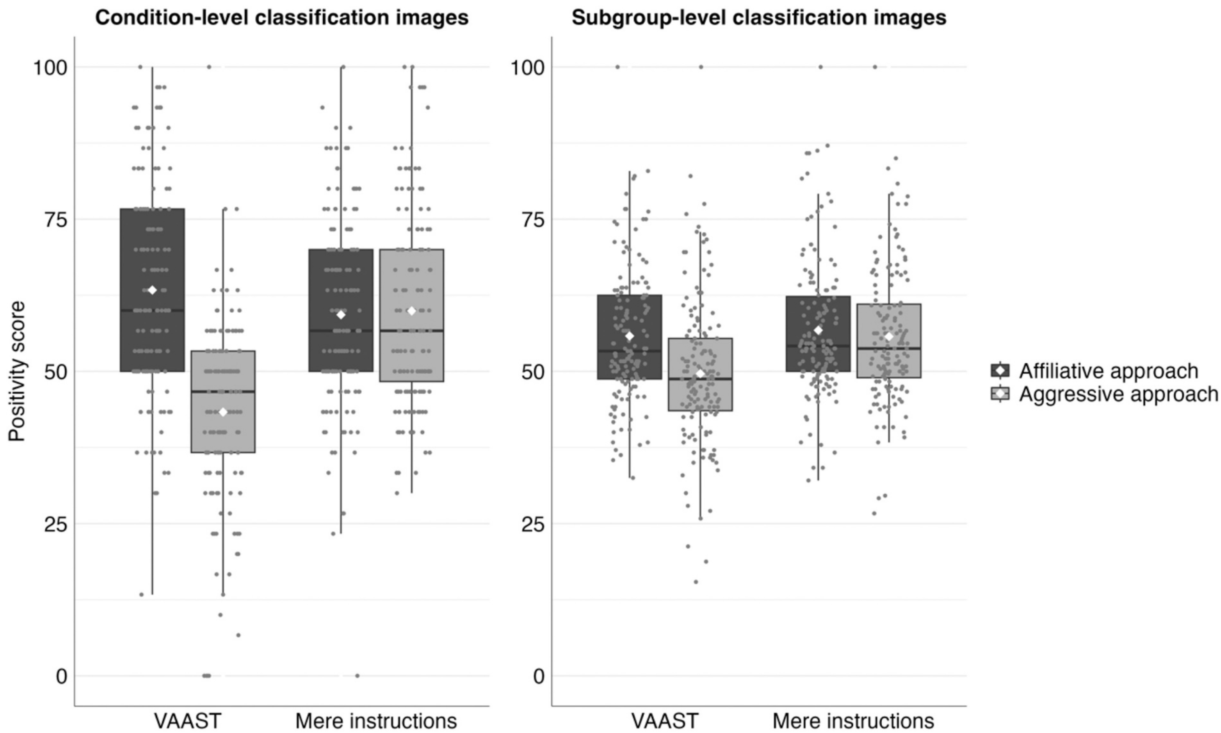

Fig. 11. Boxplot Representing the Positivity Score as a Function of the Target Action (Affiliative Approach vs. Aggressive Approach) and the Condition (VAAST vs. Mere Instuctions) for condition-level classification images (left panel) and subgroup-level classification images (right panel).
Note. The Positivity score corresponds to the average score of trustworthiness, aggressiveness, and criminality, the last two scores being reversed. Lines inside the boxes represent the median values and diamonds represent the mean values. The lower edges of boxes represent the lower quartiles and the upper edges the upper quartiles.

5. General discussion

Previous research on AAT effects provided repeated demonstrations of a "systematic" outcome of approach/avoidance, with approach leading to a more positive evaluation than avoidance. Considering that approach can sometimes take on a different meaning (e.g., aggressiveness) and convey negativity instead of positivity, we hypothesized that framing approach as aggressive (vs. affiliative) should moderate the AAT effect. Our Preliminary Experiment provided initial evidence for the framing-dependency of the AAT effect.

In Registered Experiment 1, we replicated the framing-dependency effect by manipulating actions (affiliative approach, aggressive approach, and avoidance) separately between participants. The visual representation of affiliative approach was rated as more positive than that of aggressive approach, and both were rated more positively than avoidance. Mean ratings of the visual representations of aggressive approach felt between the means of the affiliative approach and avoidance conditions. Importantly, whereas the framing effect for condition-level images was straightforward, results for subgroup-level images seemed less clear: the framing effect was significant, but the associated Bayes factor suggested anecdotal evidence. Regarding the predictions of the trait-specificity hypothesis, both visual representations of aggressive approach and avoidance were evaluated higher on avoidance-relevant traits as compared to aggressive approach relevant traits.

In Registered Experiment 2, both frequentist and Bayesian analyses were consistent with a framing effect for both condition- and subgroup-level images. We also highlighted the importance of actually experiencing approach actions during the training: the difference between affiliative and aggressive approach visual representations was larger in the VAAST condition compared to the one observed in the mere instructions condition.

5.1. Empirical contribution to approach/avoidance

The first and main contribution of this work is to provide convergent evidence for the framing-dependency of AAT effects. We found this effect in three experiments using both condition- and subgroup-level classification images – the only somewhat inconsistent result being the Bayesian analysis for subgroup images in registered Experiment 1. This finding aligns with and extends previous AAT research on the importance of framing approach/avoidance (Laham et al., 2014) and the flexible valence of these actions (Mertens et al., 2018). It also parallels studies on approach/avoidance activation suggesting that approach can be instrumental to aggression (Bossuyt et al., 2014; Krieglmeyer & Deutsch, 2013).

Second, our work raises the possibility that unframed characterization of approach may mitigate its effect. Indeed, this may leave room for inter-individual variability in the interpretation of the meaning attached to approach – if some participants interpret approach as aggressive, this should decrease the positivity of the overall approach effect. This could explain past results showing that AAT effects seem to be driven by avoidance, when compared to a control condition (Rougier et al., 2021).

Third, we did not predict a difference in negativity between visual representations of aggressive approach and avoidance – though we ventured a trait-specificity hypothesis. While the Preliminary Experiment suggested that aggressive approach leads to more negative visual representations, the effect reversed in Registered Experiment 1 when the two actions were no longer contrasted. Indeed, aggressive approach led to less negative visual representations than avoidance. This opens the possibility that, when not contrasted, approach (even when aggressive) could be associated with less negativity than avoidance to some extent. For a more definitive answer, future work could introduce a neutral control condition to inform on the general evaluative tone of aggressive approach.

Finally, the larger framing effect in the training condition compared to the instructions condition (Registered Experiment 2) suggests that the framing-dependency of AAT is not driven by demand effects (Corneille & Béna, 2023). Both conditions informed participants about the contingencies between approach vs. control actions and face groups. This implies that if participants correctly guessed the hypothesis (i.e., they were demand-aware) and tried to fake their responses (or did activate any other demand compliance process; Corneille & Lush, 2023), this should have been consistent across the two experimental conditions (i.e., instructions and VAAST). Hence, experiencing the affiliative versus aggressive approach actions, beyond any demand effects, is a likely explanation of the larger framing effect in the training condition.

5.2. Theoretical contribution

As explained in the introduction, the framing-dependency of AAT challenges some theoretical accounts (self-anchoring, motivational), while it aligns well with others (grounded cognition, common-coding, inferential). Additionally, the importance of experiencing approach actions, as compared to merely receiving instructions, challenges the inferential account because the latter predicts similar inferences about the stimuli in both conditions. This last finding, however, would fit well with a more recent, qualified version of the inferential account that differentiates past- and future-oriented inferences.

According to Rougier et al. (*in preparation*), approach/avoidance training activates past-oriented inferences (e.g., "I approached stimulus A"), while mere instructions activate future-oriented inferences (e.g., "I will approach stimulus A"). Moreover, given that past-oriented inferences refer to actual past (i.e., more certain) events, they are more likely to be used and thus to drive evaluative changes. Rougier et al. (*in preparation*) showed that participants who believed they had approached or avoided stimuli (through a bogus training and past-oriented instructions) exhibited a greater effect than those who received future-oriented instructions. Based on this qualified inferential account, the VAAST condition (promoting past-oriented inferences) is also expected to produce a larger effect compared to mere instructions (promoting future-oriented inferences). Overall, the framing-dependency of approach and the role of the training experience align with the grounded cognition approach, a qualified version of the inferential account, and the common-coding account. Future research should therefore aim to contrast more detailed predictions of these models.

First, whereas the grounded cognition approach relies on the enactment of approach/avoidance actions toward stimuli (Barsalou, 1999, 2008; Versace et al., 2014), the inferential account relies on the inferences drawn from the approach/avoidance information available, whether from actual experience or (past-oriented) instructions (Van Dessel, Hughes, & De Houwer, 2018). Thus, to disentangle these accounts, one could examine more directly the role of experiencing contingent approach/avoidance actions by comparing the effects of a real training (e.g., VAAST) with those of a bogus training (e.g., training with past-oriented instructions like "you approached stimulus A", but no actual contingencies). A larger effect from real training would support the grounded cognition approach, whereas a comparable effect between conditions would support the qualified inferential account.

Second, the grounded cognition approach predicts a pivotal role of the sensorimotor information of approach/avoidance in the observed effects because the activation of the multimodal representations in memory depends on this (see Rougier et al., 2018, 2021). On the contrary, the common-coding account does not make this prediction because codes in memory are abstracted from the (sensorimotor) information (Eder & Klauer, 2009). One could therefore manipulate the extent of sensorimotor information involved in approach/avoidance actions during the training (e.g., a rich and relevant sensorimotor information as in the VAAST vs. a more symbolic, third-person perspective condition). Contrary to the common-coding account, the grounded cognition account would predict a larger effect in the relevant sensorimotor information condition.

5.3. Limitations and future directions

The results of the current work did not confirm the trait-specificity hypothesis: visual representations of avoidance and aggressive approach were evaluated higher on avoidance-relevant traits as compared to aggressive approach-relevant traits. This pattern of findings challenges the trait-specificity hypothesis and therefore one implication of the grounded cognition account. Importantly, this discrepancy could be due to some drawbacks of the pilot study used to identify aggressive approach versus avoidance-relevant traits.

First, while we controlled for the valence and the face relevance of the traits, we did not assess the extent to which traits were equivalent on the possessor- vs. other-relevance dimension. Yet, as explained in the introduction, other-relevant facial traits (e.g., trustworthy, cruel) are more strongly related to approach/avoidance reactions than possessor-relevant facial traits (e.g., intelligent, lazy; Rougier et al., 2021; Wentura et al., 2000). Looking back at the traits we selected, it seems that avoidance-relevant traits (i.e., opportunistic, rough, dominant, racist) could be more other-relevant than approach-aggressive traits (i.e., stingy, depressive, mediocre, weak) that seem more possessor-relevant. Thus, current results may merely reflect the fact that two negatively connoted actions lead to visual representations reflecting more negative other-relevant traits than negative possessor-relevant traits.

Second, although the pilot study revealed that avoidance-relevant traits are rated as more related to avoidance than aggressive approach relevant traits, it did not show that the aggressive approach-relevant traits are rated as being more related to aggressive approach than to avoidance. In other words, we measured evaluations of avoidance and approach aggressive visual representation on more or less avoidance-relevant traits instead of avoidance-relevant and approach aggressive-relevant traits. Hence, it is possible that we did not rely on sufficiently approach aggressive-relevant traits to obtain the expected interaction between the movement and the type of trait. Combining this point with the first one, future work should investigate visual representations resulting from different actions by relying on traits clearly related to avoidance and aggressive approach and orthogonalized with other-/possessor-relevance.

Third, asking participants to identify which personality traits are associated with aggressive approach versus avoidance requires them to introspect and recall past experiences of these behaviors. However, some participants may simply lack the experience of aggressive approach (or avoidance) and thus rely on lay theories in providing their answers. Alternatively, such experiences and the features associated with aggressive approach and avoidance may vary widely between individuals, a variability that would not be reflected into the selected traits.

Fourth, the pilot did not specify whether a trait referred to the causes or the consequences of aggressive approach vs. avoidance. For instance, in case of aggressive approach, it is unclear whether participants in the pilot study generated traits associated with the person that one wants to approach to aggress (cause) or the traits of this person after the aggression (consequence). This could have led to a discrepancy between the traits generated in the pilot and the actual representation of the facial feature associated to the real actions.

Fifth, our pilot study contrasted avoidance with aggression (i.e., participants had to decide for each trait whether it was related to avoidance *or* aggression). However, Registered Experiment 1 did not directly contrast these actions, which may have reduced the likelihood of observing distinct feature patterns. Finally, the pilot study relied exclusively on negatively valenced traits, based on the assumption that aggressive approach and avoidance actions relate to negative visual representations. Yet, as discussed before, aggressive approach could be associated with some form of positivity.

Considering these limitations, future research should try to assess which facial features (both positive and negative) relate to aggression and avoidance at the individual- (rather than group-) level. These individual-level ratings could then be compared to individual-level classification images from aggressive approach vs. avoidance actions contrasted in the training. Noteworthy, a method that should facilitate the assessment of individual-level classification images is the recent Brief Reverse correlation task (Schmitz et al., 2024).

6. Conclusion

Our findings suggest that approaching stimuli can sometimes lead to more negative evaluations when approach is framed as aggressive, highlighting the importance of the meaning attached to the performed action. Moreover, merely receiving instructions about approach actions was not sufficient to produce the effect. Rather, these actions had to be effectively performed during training. Notwithstanding the steps taken in this contribution to advance our theoretical understanding, the obtained findings remain coherent with several theoretical explanations. Further delineating the merits of each of these accounts remains an important goal for future research.

Open practices

The preregistration (Preliminary Experiment, Part 2B), screenshots of power analyses (via G*Power 3), materials, data, and analytic (R) scripts are made publicly available at https://osf.io/q5f9r/?view_only=073b222de7a54b008093f1682cd6511b. Readers can also find the version of the manuscript that received the In Principe Acceptance (i.e., before data collection of Registered Experiments 1 and 2), together with the dated acceptance email from the Associate Editor (from October 31, 2023) on the repository.

CRediT authorship contribution statement

Marine Rougier: Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Mathias Schmitz:** Writing – review & editing, Writing – original draft, Software, Methodology, Conceptualization. **Ivane Nuel:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Conceptualization. **Marie-Pierre Fayant:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Baptiste Subra:** Writing – review & editing, Methodology, Conceptualization. **Theodore Alexopoulos:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Vincent Yzerbyt:** Writing – review & editing, Methodology, Conceptualization.

Declaration of competing interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2024.104697.

References

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

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences, 22*(4), 637–660. https://doi.org/10.1017/S014052 5X99532147

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Becker, D., Jostmann, N. B., Wiers, R. W., & Holland, R. W. (2015). Approach avoidance training in the eating domain: Testing the effectiveness across three single session studies. *Appetite, 85*, 58–65. https://doi.org/10.1016/j.appet.2014.11.017

Bossuyt, E., Moors, A., & De Houwer, J. (2014). On angry approach and fearful avoidance: The goal-dependent nature of emotional approach and avoidance tendencies. *Journal of Experimental Social Psychology, 50*, 118–124. https://doi.org/10.1016/j.jesp.2013.09.009

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

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision, 5*(9), 1. https://doi.org/10.1167/5.9.1

Cone, J., Brown-Iannuzzi, J. L., Lei, R., & Dotsch, R. (2021). Type I error is inflated in the two-phase reverse correlation procedure. *Social Psychological and Personality Science, 12*(5), 760–768. https://doi.org/10.1177/1948550620938616

Corneille, O., & Béna, J. (2023). Instruction-based replication studies raise challenging questions for psychological science. *Collabra: Psychology, 9*(1).

Corneille, O., & Lush, P. (2023). Sixty years after Orne's American Psychologist acticle: A conceptual framework for subjective experiences elicited by demand characteristics. *Personality and Social Psychology Review, 27*(1), 83–101. https://doi.org/10.1177/10888683221104368

Courset, R., Rougier, M., Palluel-Germain, R., Smeding, A., Jonte, J. M., Chauvin, A., & Muller, D. (2018). The Caucasian and North African French Faces (CaNAFF): A face database. *International Review of Social Psychology, 31*, 1–10. https://doi.org/10.5334/irsp.179

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition, 33*(1–2), 25–62. https://doi.org/10.1016/0010-0277(89)90005-X

Dotsch, R. (2015). *rcicr: Reverse-correlation image-classification toolbox [R package Version 0.3. 0]*.

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science, 3*, 562–571. https://doi.org/10.1177/1948550611430272

Dotsch, R., Wigboldus, D. H. J., Langner, O., & van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science, 19*, 978–980. https://doi.org/10.1111/j.1467-9280.2008.02186.x

Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (2013). Behavioral information biases the expected facial appearance of members of novel groups. *European Journal of Social Psychology, 43*, 116–125. https://doi.org/10.1002/ejsp.1928

Eder, A. B., & Klauer, K. C. (2009). A common-coding account of the bidirectional evaluation–behavior link. *Journal of Experimental Psychology: General, 138*, 218–235. https://doi.org/10.1037/a0015220

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. https://doi.org/10.3758/BF03193146

Gawronski, B., Bodenhausen, G. V., & Becker, A. P. (2007). I like it, because I like myself: Associative self-anchoring and post-decisional change of implicit evaluations. *Journal of Experimental Social Psychology, 43*(2), 221–232. https://doi.org/10.1016/j.jesp.2006.04.001

Gleibs, I. H. (2017). Are all "research fields" equal? Rethinking practice for the use of data from crowdsourcing market places. *Behavior Research Methods, 49*, 1333–1342. https://doi.org/10.3758/s13428-016-0789-y

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

Jones, C. R., Vilensky, M. R., Vasey, M. W., & Fazio, R. H. (2013). Approach behavior can mitigate predominately univalent negative attitudes: Evidence regarding insects and spiders. *Emotion, 13*, 989–996. https://doi.org/10.1037/a0033164

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2011). *Data analysis: A model comparison approach*. Routledge.

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviors. *Journal of Personality and Social Psychology, 92*, 957–971. https://doi.org/10.1037/0022-3514.92.6.957

Krieglmeyer, R., & Deutsch, R. (2013). Approach does not equal approach: Angry facial expressions evoke approach only when it serves aggression. *Social Psychological and Personality Science, 4*, 607–614. https://doi.org/10.1177/1948550612471060

Krishna, A., & Eder, A. B. (2019). The influence of pre-training evaluative responses on approach-avoidance training outcomes. *Cognition and Emotion, 33*(7), 1410–1423. https://doi.org/10.1080/02699931.2019.1568230

Laham, S. M., Kashima, Y., Dix, J., Wheeler, M., & Levis, B. (2014). Elaborated contextual framing is necessary for action-based attitude acquisition. *Cognition and Emotion, 28*, 1119–1126. https://doi.org/10.1080/02699931.2013.867833

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. https://doi.org/10.1017/cbo9781139087759

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*, 1–12. https://doi.org/10.3758/s13428-014-0458-y

Mattarozzi, K., Todorov, A., Marzocchi, M., Vicari, A., & Russo, P. M. (2015). Effects of gender and personality on first impression. *PLoS One, 10*(9), Article e0135529. https://doi.org/10.1371/journal.pone.0135529

McClelland, G. H. (2014). Nasty data: Unruly, ill-mannered observations can ruin your analysis. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 608–626). Cambridge University Press.

Mertens, G., Van Dessel, P., & De Houwer, J. (2018). The contextual malleability of approach-avoidance training effects: Approaching or avoiding fear conditioned stimuli modulates effects of approach-avoidance training. *Cognition and Emotion, 32*(2), 341–349. https://doi.org/10.1080/02699931.2017.1308315

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of Bayes factors for common designs*.

Muller, D., Bushman, B. J., Subra, B., & Ceaux, E. (2012). Are people more aggressive when they are worse off or better off than others? *Social Psychological and Personality Science, 3*(6), 754–759. https://doi.org/10.1177/1948550612436984

Neumann, R., Förster, J., & Strack, F. (2003). Motor compatibility: The bidirectional link between behavior and evaluation. In J. Musch, & K. C. Klauer (Eds.), *The psychology ofevaluation. Affective processes in cognition and emotion* (pp. 371–391). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Orne, M. T. (2009). *Demand characteristics and the concept of quasi-controls. Artifacts in behavioral research: Robert Rosenthal and Ralph L. Rosnow's classic books. 110*, 110137.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods, 46*, 1023–1031. https://doi.org/10.3758/s13428-013-0434-y

Phills, C. E., Kawakami, K., Tabi, E., Nadolny, D., & Inzlicht, M. (2011). Mind the gap: Increasing associations between the self and blacks with approach behaviors. *Journal of Personality and Social Psychology, 100*, 197–210. https://doi.org/10.1037/a0022159

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

Rinck, M., Telli, S., Kampmann, I. L., Woud, M. L., Kerstholt, M., te Velthuis, S., & Becker, E. S. (2013). Training approach-avoidance of smiling faces affects emotional vulnerability in socially anxious individuals. *Frontiers in Human Neuroscience, 7*, 1–6. https://doi.org/10.3389/fnhum.2013.00481

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237. https://doi.org/10.3758/pbr.16.2.225

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

Rougier, M., Schmitz, M., & Yzerbyt, V. (2021). When my actions shape your looks: Experience-based properties of approach/avoidance bias the visual representation of others. *Journal of Personality and Social Psychology, 120*(5), 1146–1174. https://doi.org/10.1037/pspa0000268

Rougier, M., Van Dessel, P., Moran, T., & Smith, C. T. (2024). *Can approach-avoidance instructions influence facial representations? A distinction between past- and future-oriented inferences* (in preparation).

Schmitz, M., Rougier, M., & Yzerbyt, V. (2024). Introducing the brief reverse correlation: An improved tool to assess visual representations. *European Journal of Social Psychology*. https://doi.org/10.1167/5.9.1

Schönbrodt, F. D. (2016). BFDA: Bayes factor design analysis package for R. https://github.com/nicebread/BFDA.

Schönbrodt, F. D., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review, 25*, 128–142. https://doi.org/10.3758/s13423-017-1230-y

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. https://doi.org/10.1016/j.evolhumbehav.2014.05.008

Sell, A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., & Gurven, M. (2008). Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proceedings of the Royal Society B: Biological Sciences, 276*(1656), 575–584. https://doi.org/10.1098/rspb.2008.1177

Slepian, M. L., Young, S. G., Rule, N. O., Weisbuch, M., & Ambady, N. (2012). Embodied impression formation: Social judgments and motor cues to approach and avoidance. *Social Cognition, 30*, 232–240. https://doi.org/10.1521/soco.2012.30.2.232

Smith, C. T., Calanchini, J., Hughes, S., Van Dessel, P., & De Houwer, J. (2019). The impact of instruction- and experience-based evaluative learning on IAT performance: A Quad model perspective. *Cognition and Emotion, 34*, 21–41. https://doi.org/10.1080/02699931.2019.1592118

Solarz, A. K. (1960). Latency of instrumental responses as a function of compatibility with the meaning of eliciting verbal signs. *Journal of Experimental Psychology, 59*, 239–245. https://doi.org/10.1037/h0047274

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*, 1096–1104. https://doi.org/10.3758/brm.42.4.1096

Stoet, G. (2017). A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology, 44*, 24–31. https://doi.org/10.1177/0098628316677643

Sutherland, C. A. M., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial first impressions across culture: Data-driven modeling of Chinese and British perceivers' unconstrained facial impressions. *Personality and Social Psychology Bulletin, 44*(4), 521–537. https://doi.org/10.1177/0146167217744194

Van Dessel, P., De Houwer, J., Gast, A., Roets, A., & Smith, C. T. (2020). On the effectiveness of approach-avoidance instructions and training for changing evaluations of social groups. *Journal of Personality and Social Psychology, 119*(2), e1–e14. https://doi.org/10.1037/pspa0000189

Van Dessel, P., De Houwer, J., Gast, A., & Smith, C. T. (2015). Instruction-based approach-avoidance effects. *Experimental Psychology, 62*, 161–169. https://doi.org/10.1027/1618-3169/a000282

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

Van Dessel, P., Eder, A. B., & Hughes, S. (2018). Mechanisms underlying effects of approach-avoidance training on stimulus evaluation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(8), 1224–1241. https://doi.org/10.1037/xlm0000514

Van Dessel, P., Hughes, S., & De Houwer, J. (2018). How do actions influence attitudes? An inferential account of the impact of action performance on stimulus evaluation. *Personality and Social Psychology Review, 23*, 267–284. https://doi.org/10.1177/1088868318795730

Versace, R., Vallet, G. T., Riou, B., Lesourd, M., Labeye, É., & Brunel, L. (2014). Act-in: An integrated view of memory mechanisms. *Journal of Cognitive Psychology, 26*(3), 280–306. https://doi.org/10.1080/20445911.2014.892113

Wang, H., Han, C., Hahn, A. C., Fasolt, V., Morrison, D. K., Holzleitner, I. J., … Jones, B. C. (2019). A data-driven study of Chinese participants' social judgments of Chinese faces. *PLoS One, 14*(1), Article e0210315. https://doi.org/10.1371/journal.pone.0210315

Wentura, D., Rothermund, K., & Bak, P. (2000). Automatic vigilance: The attention-grabbing power of approach- and avoidance-related social information. *Journal of Personality and Social Psychology, 78*, 1024–1037. https://doi.org/10.1037/0022-3514.78.6.1024

Wiers, R. W., Eberl, C., Rinck, M., Becker, E. S., & Lindenmeyer, J. (2011). Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science, 22*(4), 490–497. https://doi.org/10.1177/0956797611400615

Wiers, R. W., Rinck, M., Kordts, R., Houben, K., & Strack, F. (2010). Retraining automatic action-tendencies to approach alcohol in hazardous drinkers. *Addiction, 105*, 279–287. https://doi.org/10.1111/j.1360-0443.2009.02775.x

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750

Woud, M. L., Becker, E. S., Lange, W.-G., & Rinck, M. (2013). Effects of approach-avoidance training on implicit and explicit evaluations of neutral, angry, and smiling face stimuli. *Psychological Reports, 113*, 199–216. https://doi.org/10.2466/21.07.pr0.113x10z1

Zebrowitz, L. A., Fellous, J. M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. *Personality and Social Psychology Review, 7*(3), 194–215. https://doi.org/10.1207/S15327957PSPR0703_01