

Fear Recognition in Healthy and Clinical Adults: A Systematic Review and Meta-analysis

A. T. Martins

University of Algarve

F. R. Carvalho

frcarvalho@ualg.pt

University of Algarve

J. Russo

University of Algarve

C. Campos

University of Algarve

J.A. Camacho-Ruiz

Department of Personality, Evaluation and Psychological Treatment

C.M. Galvez-Sánchez

Universidad de Murcia

J. Rodrigues

University of Algarve

A. Gomes

University of Algarve

Article

Keywords: Fear Facial Recognition, Systematic Review with Meta-Analysis, Fear- Surprise confusion

Posted Date: August 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10347133/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Facial fear recognition is a behavioural marker of theoretical and clinical relevance, but the literature is fragmented by methodological and cultural heterogeneity. This PRISMA 2020 systematic review and meta-analysis synthesises 40 studies (2016–2025) on fear recognition in healthy and clinical adults. Risk of bias was assessed with a hybrid PROBAST/AXIS tool. Pooled fear accuracy in healthy controls, from a DerSimonian–Laird random-effects logit model across 21 studies ($n = 1369$), was 54.5% (95% CI 45.8–62.9), with very large heterogeneity ($I^2 = 87.6\%$) and a range of 19% to 94%. This dispersion was largely methodological: paradigms with fewer than six response options yielded a far higher raw pool (80.7%; $k = 4$) than those with six or more (47.8%; $k = 17$), reflecting the higher chance baseline and the exclusion of surprise rather than a true difference in ability. Across three studies with complete confusion matrices, the pooled fear $\rightarrow$ surprise rate was 32.3% (95% CI 25.5–39.8; $I^2 = 4\%$), conserved across method and culture. Aggregate estimates should be read alongside subgroup analyses. The fear–surprise confusion may be a differential indicator worth assessing with standardized procedures and detailed error analyses.

INTRODUCTION

Facial emotion recognition as a fundamental sociocognitive capacity

Facial emotion recognition (FER) is one of the most extensively studied sociocognitive abilities of the past six decades ¹. Building on Darwin's thesis on the universality of emotional expressions² and on the empirical demonstration of cross-cultural recognition of six basic emotions³, the literature long treated the face as a primary channel of interpersonal affective communication ^{4,5}, a position that has since been actively debated (see below). The Facial Action Coding System (FACS) operationalized this insight by decomposing every emotional expression into an identifiable set of action units (AUs) controlled by specific facial muscles ³, offering an anatomical foundation for rigorous FER research ⁶. The capacity to correctly recognize facial emotional expressions is implicated in multiple domains of everyday life, including social adjustment, emotion regulation, the formation of close relationships, and threat detection ⁷. Acquired FER deficits in adulthood are associated with functional deterioration in professional and family domains ^{8,9}, and FER has become a privileged behavioral marker in both experimental and clinical neuropsychology ¹⁰. Its measurement has, however, been criticized for methodological inconsistency ¹¹. This review focuses on fear, the basic emotion for which such inconsistency is most consequential, and quantifies how method, culture and clinical status jointly shape recognition ¹².

Although this framework has guided six decades of research, its central premise – that a discrete set of facial configurations maps reliably and specifically onto basic-emotion categories across cultures – has been substantially contested. Methodological reviews have argued that the high recognition rates

reported in classic studies are inflated by forced-choice response formats, within-subject designs, and the use of pre-selected, prototypical stimuli ¹. Constructivist accounts further propose that the correspondence between facial movements and emotion categories is weak, variable, and strongly shaped by context and conceptual knowledge, rather than constituting a fixed emotional "read-out" ^{13,14}. Cross-cultural work using reverse-correlation and free-labelling methods has likewise shown that observers from different cultures do not represent the six emotions with equivalent facial signals, with the overlap being most pronounced among the negatively valenced, high-arousal expressions. At the same time, large-scale naturalistic studies continue to report cross-cultural regularities in the contexts in which specific expressions occur ^{15,16}, and integrative proposals increasingly treat the categorical and constructivist positions as complementary rather than mutually exclusive ^{17,18}. The present review does not adjudicate this debate a priori; instead, it treats the accuracy and confusion structure of fear recognition as empirical evidence that bears on it.

Fear as a special case: theoretical and clinical relevance

Among the six basic emotions, fear occupies a special position ⁴. From an evolutionary standpoint, fear is the emotion that signals imminent threat and recognizing it in a conspecific's face is an adaptive capacity for socially mediated threat detection ⁷. Accordingly, fear is the emotion most consistently associated with rapid autonomic responses and with preferential activation of the bilateral amygdala ⁴. Bilateral amygdala lesions produce selective deficits in the recognition of fear while sparing the recognition of other emotions ^{4,19}. Deficits in fear recognition are consistently reported in multiple clinical conditions, including Parkinson's disease^{9,20} and amyotrophic lateral sclerosis ²¹. Comparable patterns have been described in fronto-temporal dementias^{22,23} and in mesial temporal lobe epilepsy ²⁴. In psychiatric populations, selective deficits in fear recognition emerge in major depressive disorder ²⁵, schizophrenia²⁶ and obsessive-compulsive disorder ²⁷. This wide range suggests that fear recognition is particularly vulnerable to multiple types of brain and psychiatric pathology ¹⁰, making it a potentially sensitive marker in clinical settings ²⁸. It is important, however, to distinguish two senses in which fear is "special". First, it is an emotion biologically prepared to signal threat with neural salience ⁴. Second, and paradoxically, it is the most difficult facial emotion to recognize under standardized conditions ¹, difficulty emerges with regularity both in healthy controls²⁹ and in clinical samples ²¹. Across different samples and paradigms, fear is systematically confused with surprise ^{6,19}, owing to the overlap of their action units, and this perceptual difficulty is replicated in signal-detection and threshold paradigms ²⁵. These two properties are often presented as paradoxical, yet they are reconcilable within an account in which emotional signals are transmitted as an evolving hierarchy over time ¹⁸. On this view, fear and surprise share an early, biologically salient signal, principally the widening of the eyes (AU1, AU2, AU5), that is transmitted before the lower-face actions (e.g., AU20, AU25) that subsequently disambiguate the two expressions. Neural salience and perceptual fragility may thus reflect one structure: a signal rapidly flagged as threat-relevant but slowly resolved into a category. Consistent with this, the fear-recognition deficit following amygdala damage has been attributed not to a loss of the fear concept as such, but to a

failure to spontaneously fixate the eye region, a deficit that is reversed when attention is explicitly directed there³⁰. The dual nature of fear, biologically salient yet perceptually fragile, thus motivates a careful systematic quantification of its accuracy and confusion patterns across the existing literature¹², and renders the direction of its misclassification, not merely its accuracy, a behavioral window onto how the perceptual system weights shared versus distinctive facial cues

Heterogeneity in literature and the case for systematic synthesis

Despite over five decades of FER research, the literature remains fragmented by methodological heterogeneity^{1,11}. Paradigms range from six- or seven-option forced choice to pairwise 2-AFC²⁵ and oddball ERP designs³¹. Stimuli vary between static photographs³², dynamic morphs^{6,26}, and multimodal videos^{33,34}. Cultural heterogeneity is also substantial^{35,36}. A third layer is clinical: studies with healthy controls only³⁷ and with clinical comparisons³⁸ coexist. Without rigorous quantitative synthesis, cross-study comparisons are vulnerable to comparability problems³⁹, and clinical interpretation lacks calibrated reference points⁹. Three recent developments justify an updated synthesis¹²: the broad introduction of dynamic stimuli^{6,20}; the expansion of multimodal paradigms^{29,33}; and the COVID-19 pandemic, which introduced facial masks as a pervasive ecological variable^{40–43}.

Prior reviews and the gap addressed here

Prior reviews and meta-analyses have addressed FER in specific clinical conditions, such as Parkinson's disease⁹, schizophrenia¹⁰, and dementias^{8,22}. Cross-cultural FER reviews³⁵ and reviews of the mask effect^{40,43} also exist. To our knowledge, however, no systematic review has combined the three distinctive features pursued here. First, an explicit focus on fear as the target emotion^{4,24}. Second, the simultaneous inclusion of healthy controls^{29,37} and clinical populations^{21,27}. Third, a quantitative synthesis of both accuracy and cross-emotion confusion patterns^{25,44}, combined with a formal risk-of-bias assessment^{45,46}. The present review addresses these three gaps¹².

Objectives

This systematic review was conducted under scientific supervision and follows the PRISMA 2020 reporting guideline¹². The general objective of this study was to conduct a systematic review and meta-analysis to synthesize the evidence on fear facial recognition in healthy adults and clinical populations, assessing accuracy, confusion patterns, and the methodological and clinical factors influencing these outcomes. The specific objectives are five:

1. To quantify facial recognition of fear in healthy adults ¹. We compute a random-effects logit pool of fear accuracy in HC across studies published between 2016 and 2025 ^{47,48}.
2. To identify methodological moderators of fear accuracy ³⁹. We examine the effect of modality, culture ³⁶, era of publication ⁴⁰ and number of response options.
3. To quantify the pattern of confusion between fear and other basic emotions ¹⁹. We compute pools for fear → other-emotion confusions in studies with complete confusion matrices ^{29,44,49}.
4. To narratively synthesize alterations in clinical populations ⁹. We describe fear recognition in depression ²⁵, schizophrenia ²⁶, Parkinson's disease ²⁰, and PTSD ²⁸.
5. To assess methodological quality and risk of bias ⁴⁵. We apply a hybrid PROBAST/AXIS tool adapted to FER studies ⁴⁶ and derive a continuous Quality Score for each study.

Hypotheses

Based on prior literature, we pre-specified three hypotheses ³⁹:

H1. Fear accuracy in healthy controls varies substantially across studies ($I^2 > 75\%$), reflecting methodological and cultural heterogeneity ^{1,32,36}.

H2. When fear is misclassified in healthy controls, it is more frequently confused with surprise than with any other emotion ^{19,29}, and this direction of confusion is stable across studies ($I^2 < 25\%$ for the fear → surprise pool; ^{25,44}).

H3. Paradigms with six or more response options yield substantially lower fear accuracy pools than paradigms with fewer than six options, due to the opportunity for the fear → surprise confusion ^{25,40}.

Detailed results testing these hypotheses are presented in Section 3, with interpretation in Section 4.

METHODS

RESULTS

Study selection

The database search identified 2,835 records (302 in PubMed/MEDLINE and 2,533 in Scopus). After removal of 126 duplicates, 2,709 records were screened at title and abstract, of which 2,663 were excluded, leaving 46 records that were sought for full-text retrieval. Of these, two could not be retrieved ^{44,50} (full-text unavailable through institutional or open-access channels). The remaining 44 records underwent full-text eligibility assessment. Four records were excluded at this stage according to pre-specified criteria: ⁵¹ for inadequate population (children/adolescents aged 6–16 years), ⁵² and ⁵³ for exclusively clinical samples (bipolar disorder and treatment-resistant schizophrenia, respectively)

without healthy control groups, and one further record⁵⁴ was excluded for the absence of any fear-specific measurement (a content criterion confirmable only at full text). The final review therefore includes 40 studies published between 2016 and 2025 (Fig. 1). The PRISMA 2020 flow diagram (Fig. 1) summarizes the identification, screening, eligibility and inclusion phases, in accordance with¹². Per-study characteristics of the 40 included studies are reported in Supplementary Table S3; the full list of excluded studies with reasons is in Supplementary Table S4; the search strategy details are in the Supplementary Methods; the PRISMA 2020 main and abstract checklists are in Supplementary Tables S5 and S6.

Characteristics of included studies

The 40 included studies were published between 2016 and 2025, with 14 (35%) in or after 2020 (the COVID-19 era). Studies were predominantly European ($n = 13$) and Asian ($n = 11$), followed by North American ($n = 8$), South American ($n = 4$) and African ($n = 1$). The cumulative sample is approximately 2400 participants; the meta-analytic pool of healthy controls (Section 3.4) includes 1369 from 21 studies. Eighteen studies assessed exclusively healthy controls and five included a non-clinical at-risk or individual-differences sample (attachment style, pandemic confinement, subclinical depressive symptoms, interoceptive awareness, mask exposure). The remaining seventeen included a clinical group alongside healthy controls: schizophrenia ($n = 3$), major depressive disorder ($n = 3$), Parkinson's disease ($n = 2$), amyotrophic lateral sclerosis ($n = 1$) and forensic/offender samples ($n = 2$), with single studies of bipolar disorder, post-traumatic stress disorder, ADHD, mesial temporal lobe epilepsy, amnesic mild cognitive impairment, and alcohol use, among others. The predominance of single-study conditions precludes condition-level meta-analytic aggregation and is itself a marker of the literature's fragmentation. Stimuli were static photographs in 25 studies (62.5%), dynamic morphs/videos in 11 (27.5%), and multimodal in 4 (10%). Tasks were six- or seven-option forced choice ($n = 17$), pairwise 2-AFC ($n = 2$), oddball or go/no-go ($n = 4$), Likert intensity ratings ($n = 7$), and specialized paradigms (eye-tracking, QUEST, ERP/EEG, bodily sensation maps; $n = 10$).

Risk of bias

The hybrid PROBAST/AXIS tool was applied to all 40 included studies, yielding 0 studies with LOW overall classification, 33 (82.5%) UNCLEAR, and 7 (17.5%) HIGH (Supplementary Fig. S1). This distribution mostly reflects the omission of formal sample-size justification, complete stimulus description, and explicit handling of missing data — three items that are rarely reported with the level of detail required for LOW. The continuous Quality Score across all 40 included studies had a median of 65% (range 7%–93%); the median within the 21 studies contributing to the meta-analytic pool was higher (71%, see § 3.7). The five studies with the highest Quality Scores were²⁷ 93%,²⁹ 93%,²⁵ 86%,⁵⁵ 86%, and⁵⁶ 86%. The five lowest were⁵⁷ 7%,⁵⁸ 7%, both without retrievable PDF),⁵⁹ 17%,⁶⁰ 36%, and⁴⁹ 36%, mostly because of incomplete method reporting. Domain D2 (Stimuli) was the most polarized, accumulating both the highest number of HIGH classifications ($n = 5$; paradigms with limited information on stimulus intensity, model numbers, or prior validation) and the highest number of LOW classifications ($n = 11$;

studies using fully validated facial-stimulus databases). By contrast, Domain D1 (Participants) had the fewest LOW classifications (n = 4), because sample representativeness was rated unclear in most studies (30 of 40). At the item level, however, the majority of the 560 individual judgements were low risk (343/560, 61.2%; 200, 35.7% unclear; 7, 1.2% high; 10 not applicable); the predominance of UNCLEAR overall classifications is therefore a conservative artefact of the domain-level rule. The complete study × item heatmap is shown in Supplementary Fig. S1.

Meta-analytic pool of fear accuracy in healthy controls

Of the 40 included studies, 21 reported quantifiable fear accuracy in healthy controls and were included in the meta-analytic pool (Section 2.7). The pool was computed using the DerSimonian–Laird random-effects model with a logit transformation. Table 1 (and Fig. 2) summarizes the results.

Table 1	
Primary meta-analytic pool (random-effects logit; k = 21).	
Metric	Value
k (studies in pool)	21
Total n (healthy controls)	1369
Pooled estimate (random-effects, logit)	54.5%
95% CI	45.8–62.9
I ² heterogeneity	87.6%
τ ² (logit scale)	0.527
Cochran's Q (df = 20)	160.94, p < .0001

Twenty-one studies contributed to the pool (full per-study characteristics in Supplementary Table S3). The pooled estimate of fear accuracy in healthy controls is 54.5% (95% CI 45.8–62.9), with extremely high heterogeneity (I² = 87.6%; τ² = 0.527; Cochran's Q = 160.94, df = 20, p < .0001; ³⁹). The observed range spans from 19.4% ³² to 94.3% ⁴², a difference of 74.9 percentage points. This level of heterogeneity implies that the aggregated pool should be interpreted with caution; the subgroup analyses (Section 3.5) are more informative.

3.5. Subgroup analyses

We computed five pre-specified subgroup analyses to investigate moderators of fear accuracy: stimulus modality, cultural background, era of publication, and number of response options. Table 2 (and Fig. 3) summarizes the pools by subgroup.

Table 2
Subgroup pools (random-effects logit; see also Fig. 3). Subgroup differences were driven chiefly by the number of response options, a largely methodological effect rather than a true moderator of ability.

Subgroup	k	n_HC	Pool % [95% CI]	I ²
All studies	21	1369	54.5 [45.8–62.9]	88%
Modality: static	14	925	56.7 [45.8–66.9]	88%
Modality: dynamic	6	223	53.2 [35.0–70.6]	83%
Culture: Western	11	731	57.3 [46.5–67.5]	84%
Culture: East Asian	9	573	52.1 [35.6–68.2]	91%
Year ≤ 2019	7	466	53.5 [42.1–64.5]	78%
Year ≥ 2020	14	903	55.2 [42.9–66.9]	90%
Response options < 6	4	177	80.7 [68.2–89.1]	60%
Response options ≥ 6	17	1192	47.8 [39.5–56.1]	85%

The largest subgroup difference was the number of response options: pools with six or more options ($k = 17$) yielded 47.8% [39.5–56.1] versus 80.7% [68.2–89.1] for those with fewer than six ($k = 4$) — a 33-point gap with non-overlapping CIs. This is largely methodological: six-plus-option paradigms typically include surprise, creating the fear → surprise confusion (Section 3.6). Other moderators were smaller and less robust: modality (56.7% vs 53.2%), culture (57.3% vs 52.1%), and pandemic era (53.5% vs 55.2%), all with overlapping or near-identical estimates. High within-subgroup heterogeneity (I^2 60–91%) points to further moderators not captured by the pre-specified variables (e.g., stimulus intensity, model identity, mask exposure).

3.6. Cross-emotion confusion patterns

Three studies published complete fear confusion matrices for healthy controls in the main paper: Puffet et al. (2025; $n = 76$, 7-option static, ECIFBSS)²⁹, Calvo et al. (2018; $n = 72$, 6-option dynamic, KDEF-dyn)⁴⁴ and Kosonogov & Titova (2019; $n \approx 25$, 7-option static)⁴⁹. The distribution of categorizations for fear stimuli across these three studies is shown in Fig. 4 (panel A). Random-effects pools for each fear → X pair were computed (Table 3; Fig. 4 panels B and C).

Table 3

Pooled rates of fear → other-emotion confusions in healthy controls (random-effects logit). Surprise is the predominant misclassification across the available studies. Note. Pools derive from the three studies that report fear confusion data for healthy controls (Puffet 2025, $n = 76$; Calvo 2018, $n = 72$; Kosonogov 2019, $n \approx 25$). All three report the fear → surprise cell ($k = 3$); Kosonogov reported only the surprise confusion, so the remaining cells pool the two studies with complete matrices ($k = 2$). The listed cells are off-diagonal confusions only and do not sum to 100%, as they exclude correct responses and derive from partially overlapping study subsets. Fear was correctly recognised in roughly half of trials in these studies (49.5%, 59.8% and 40.6%, respectively). A fear → neutral confusion was reported by a single study (Puffet, 2.66%) and was therefore not pooled. Given $k = 2-3$, the I^2 values and confidence intervals for the non-surprise cells should be interpreted with caution.

Confusion	k	Pool % [95% CI]	I^2
fear → surprise	3	32.3 [25.5–39.8]	4%
fear → disgust	2	8.3 [4.5–14.7]	14%
fear → sadness	2	2.9 [1.1–7.3]	0%
fear → anger	2	1.4 [0.4–5.6]	0%
fear → happiness	2	0.4 [0.0–5.1]	0%

The fear → surprise pool was 32.3% (95% CI 25.5–39.8) with minimal heterogeneity ($I^2 = 4\%$; $\tau^2 \approx 0$), a striking contrast with the $I^2 \approx 88\%$ of the overall accuracy pool³⁹. This stability, despite the wide absolute-accuracy variability, indicates that the fear–surprise equivalence is conserved across method and culture. The other targets were all below 9% and compatible with zero, except fear → disgust.

Across 22 studies, 9 reported a dominant confusion direction, of which 8 (89%) named surprise^{6,24–26,29,41,49,61} and only 1 named sadness⁶² (in audio-only modality). The remaining 13 studies did not report a specific direction ($n = 11$) or used paradigms incompatible with confusion assessment ($n = 2$).

Sensitivity analyses

To assess publication bias, we constructed a funnel plot of standard error against logit-transformed fear accuracy (Supplementary Fig. S2). Egger's regression⁶³ yielded an intercept of 2.02 ($p = 0.196$), not statistically significant, indicating no detectable funnel asymmetry. Two caveats apply: $k = 21$ is only marginally above the recommended minimum for Egger's test, and the test has limited power against subtle asymmetry. Two further sensitivity analyses were therefore conducted (Supplementary Fig. S3). A leave-one-out analysis yielded pooled values ranging from 52.6% to 56.4% — a maximum shift of 3.8 percentage points around the full pool of 54.5%, indicating that no single study disproportionately drives

the estimate. The most influential studies^{25,32,41,42} are those with the most extreme fear-accuracy values. A second pool, restricted to studies with a Quality Score above the pool median ($k = 12$; $QS \geq 71\%$), yielded 59.3% (95% CI 48.5–69.3, $I^2 = 86\%$), slightly higher than the full pool but within its confidence interval and equally heterogeneous. The pool is therefore robust to both.

Narrative synthesis of clinical populations

Beyond the meta-analytic pool restricted to healthy controls, we narratively synthesised the 33 studies that included a clinical comparison and those that focused on individual differences in non-categorical paradigms (e.g., trustworthiness perception;⁶⁴). Four emergent patterns deserve mention. First, fear recognition deficits are consistently reported in major depressive disorder^{20,25} by meta-analysis, Parkinson's disease^{20,21}, and amyotrophic lateral sclerosis²¹. Second, in PTSD²⁸, the bias is qualitatively different: it is not fear that is misrecognised, but rather neutral expressions that are interpreted as threatening (anger, fear, disgust), the so-called "threat interpretation bias". Third, in some disorders, dynamic stimuli attenuate or eliminate deficits observed with static stimuli^{20,26} in part, for schizophrenia, suggesting that the ecological validity of the stimulus moderates the expression of the deficit. Fourth, the direction of confusion may differ from the healthy baseline: in schizophrenia²⁶, fear is more frequently confused with disgust (16.7%) than with surprise, indicating a pathological reorganization of the emotional perceptual space; in MDD²⁵, the dominant fear $\leftrightarrow$ surprise pattern is preserved but the magnitude of the error increases.

Summary of main findings

1. Primary fear-accuracy pool in healthy controls: 54.5% [95% CI 45.8–62.9], $k = 21$, $n = 1369$, with very high heterogeneity ($I^2 = 88\%$; $\tau^2 = 0.53$) reflecting an observed range from 19% to 94%.
2. The dominant moderator is the number of response options: pools with ≥ 6 options (47.8%) differ from pools with < 6 options (80.7%) by more than 30 percentage points.
3. Dominant confusion direction: fear $\rightarrow$ surprise pool 32.3% [25.5–39.8], $I^2 = 4\%$ ($k = 3$ studies with full confusion matrices) – stable across studies despite the huge variation in absolute accuracy.
4. Categorical synthesis: 8 of 9 studies reporting a specific confusion identified surprise as the dominant confusion of fear in healthy controls (89%).
5. Risk of bias: 0 LOW overall, 33 UNCLEAR, 7 HIGH; mean Quality Score 62% (range 7%–93%).
6. No detectable publication bias (Egger's intercept = 2.02, $p = 0.196$).

DISCUSSION

Fear recognition shapes social interaction and emotional regulation, and its confusion patterns bear on psychiatric and neurological diagnosis. From our synthesis, three key findings warrant discussion. First, fear accuracy in healthy controls varies massively, from 19%³² to 94%⁴², yielding a pool with $I^2 \approx 88\%$ ³⁹. This 75-point amplitude reflects systematic differences in method^{1,40}, culture³⁶ and stimuli^{26,33}. The aggregate pool (54.5%; 95% CI 45.8–62.9; $k = 21$) should therefore be read with caution⁴⁷ and is

subordinated to the subgroup analyses. Second, despite this variability, the direction of misclassification is conserved: fear is categorized as surprise in roughly one-third of cases (pool fear surprise = 32.3%; $I^2 \approx 4\%$; ^{29,44}. This stability suggests that fear surprise perceptual equivalence is a robust phenomenon ^{3,19}. Third, fear remains the most difficult facial emotion to recognize among the six basic emotions ^{1,11}, even in healthy samples ^{29,37}. In at least 20 of 40 studies, fear had the lowest or second-lowest accuracy; in d $\bar{x}$ studies ^{25,26}, it had the lowest signal-detection sensitivity. The perceptual vulnerability of fear is thus a consensus finding ⁴.

The largest source of between-study heterogeneity was methodological rather than substantive – the number of response options ²⁵: paradigms with six or seven options produced much lower pools (47.8%; $k = 17$) than paradigms with fewer than six options (80.7%; $k = 4$). This 33-percentage-point difference reflects predominantly the fact that paradigms with six or more options typically include surprise as an option, creating the opportunity for the fear $\rightarrow$ surprise error ^{6,19}. Two- or three-option forced-choice paradigms, or pairwise designs ²⁵, eliminate this error source by design and also raise the chance-level baseline, so that raw accuracy is not directly comparable across response-set sizes; this low-option pool rests on only four studies ($n = 177$) and should be read as a demonstration of this methodological artefact, not as evidence of superior fear recognition. A second important moderator was stimulus dynamism: dynamic stimuli (morphs, videos) yielded higher pools than static stimuli ^{20,26,65}, and several clinical studies report that dynamic stimuli attenuate deficits observed with static stimuli ²⁰ in PD and for schizophrenia²⁶, with therapeutic implications ⁹. Culture emerged as a third moderator, although confounded with method ³⁶: Asian samples produced both the lowest ³² and the highest ²⁵ values of the pool. The literature on cultural emotional dialects ³⁵ suggests that East Asians tend to have lower fear accuracy with Western Ekman stimuli, but the paradigm effect is dominant ^{61,66}. The pandemic era (2020+) introduced new heterogeneity ^{40,43}. Studies such as Yamasaki & Miyata ⁴¹ and Abutalebi ⁴² measured the explicit effect of facial masks on fear recognition, showing massive reductions under masked conditions. These findings suggest that post-2020 literature must be interpreted with mask exposure as an additional ecological variable.

The most robust finding of this meta-analysis is the stability of the fear $\rightarrow$ surprise confusion across healthy-control samples ^{29,44,49}. Three studies with complete confusion matrices produced a pool of 32.3% (95% CI 25.5–39.8) with near-zero heterogeneity ($I^2 = 4\%$; $\tau^2 \approx 0$; ³⁹). The phenomenological interpretation is direct: fear and surprise share critical facial action units, AU1 (inner brow raise), AU2 (outer brow raise), and AU5 (upper lid raise) ^{3,40}. The two expressions also differ in lower-face action units, fear typically includes AU20 (lip stretcher) and AU25 (lips part), which surprise lacks. The persistent confusion despite these distinguishing lower-face cues suggests that observer's underweight the lower face when identifying fear, attending preferentially to the shared upper-face configuration. This structural overlap, well described in the FACS literature ⁶⁷, has been characterized dynamically by reverse-correlation models showing that the shared upper-face signal is transmitted early, before the lower-face actions that later disambiguate the two expressions ⁶⁸. The near-zero heterogeneity of the

fear → surprise pool across methods and cultures, despite the wide variability in absolute accuracy, is consistent with this account: a small set of biologically rooted signals is shared early in transmission and differentiated later by culturally and contextually modulated cues^{18,68}. This convergence spans markedly different methods, forced-choice behavioral pooling here, data-driven signal reconstruction there (Jack, Garrod, Yu, Caldara, & Schyns, 2012), strengthening the inference that this equivalence is conserved rather than a single-paradigm artefact. In addition, 8 of the 9 studies that classified a specific confusion identified surprise as the dominant confusion of fear^{24,25,41,61}. Yamasaki & Miyata⁴¹ report a Cohen's $d = 1.39$ for the fear surprise misclassification under masked conditions, a very large effect size. Iannaccone²⁴ further note that fear and surprise uniquely require attention to both the eyes and the mouth, unlike happiness (mouth-dominated) and anger/sadness (eye-dominated;^{37,44}). Three implications follow. First, the "difficulty" of fear is not intrinsic to the emotion¹ but a structural property of the perceptual space^{3,19}. Second, clinical comparisons of fear accuracy between groups must control for the response paradigm^{11,25}: a 10-point difference is trivial in a six-option paradigm but substantial in a pairwise design. Third, training interventions for fear recognition may benefit from paradigms explicitly targeting fear-vs-surprise discrimination⁶⁹.

Methodological implications for future research

Four concrete recommendations follow. First, raw accuracy is an inadequate metric for cross-cultural or cross-paradigm comparisons; it should be replaced or complemented with the unbiased hit rate, which adjusts for response bias^{11,29}. Second, reporting the complete confusion matrix should be a norm. Of the 40 included studies, only three published a complete 6- or 7-emotion confusion matrix for healthy controls^{29,44,49}. Mo et al.²⁵ published a complementary pairwise F1 matrix but not a forced-choice category matrix. This shortage severely limits the power of meta-analyses to test confusion-specific hypotheses. Third, seven-option paradigms (with neutral) are ecologically more valid than six-option paradigms; researchers should report results under both. Fourth, given the heterogeneity of the global pool, future meta-analyses should prioritize subgroup analyses over aggregate pools^{39,47}.

Clinical implications: using fear as a marker

The perceptual vulnerability of fear¹, combined with its directional stability⁴⁴, makes it a sensitive marker in several clinical contexts. In limbic-predominant amnesic mild cognitive impairment^{8,22,24}, fear recognition is selectively impaired relative to the amnesic Alzheimer phenotype; this pattern is consistent with the dependence of fear recognition on the amygdala and insula⁷⁰. In schizophrenia, fear shows both accuracy deficits and systematic confusion with disgust or anger, diverging from the healthy baseline^{10,26}. In MDD^{25,71}, Parkinson's disease^{9,20} and ALS²¹, the deficit is consistent. Three clinical implications emerge from these patterns; however, all rest on limited single-study evidence and require independent replication. First, we tentatively describe a candidate "fear non-surprise" profile — a confusion-matrix pattern in which the most frequent misclassification of fear is not surprise but rather disgust, anger, sadness, neutral or happiness. Because the present review shows that the fear →

surprise rate is itself strongly shaped by the response-option set, this profile cannot yet be attributed to pathological reorganization of the emotional perceptual space; observed so far in a single exploratory dataset (²⁶, in schizophrenia; $k = 1$), it should be regarded as a hypothesis for future research, to be tested only after the response-option set is held constant. Second, in PTSD ²⁸, the bias is different, it is not fear that is misrecognized but rather neutral expressions that are interpreted as fear/anger/disgust ("threat interpretation bias"). Third, the dynamism of the stimulus may normalize deficits in some disorders ²⁰, suggesting that interventions and diagnostic assessments should use ecologically valid stimuli whenever possible ^{26,33}.

Limitations

The present analysis is subject to several important limitations. First, global pooled heterogeneity was very high ($I^2 = 88\%$; $\tau^2 = 0.53$), which substantially constrains the interpretability of the aggregate estimate. This level of variability suggests that the included studies differ considerably beyond chance, and therefore conclusions should be drawn with caution.

Second, complete confusion matrices for healthy controls were available in only 3 out of 40 studies, rendering the fear–surprise confusion pool ($k = 3$) underpowered. With such a small number of studies, the I^2 estimate is highly uncertain. Consequently, the observed near-zero heterogeneity should be interpreted as an indication of consistency across the limited available studies, rather than as a precise estimate of population-level variability.

A further limitation concerns potential selection-bias circularity within the confusion pool. All three studies contributing to the fear → surprise analysis used 6- or 7-option forced-choice paradigms in which “surprise” was explicitly included as a response option. In contrast, paradigms that do not include surprise (e.g., pairwise two alternative forced-choice tasks) cannot, by design, detect this type of confusion. As such, the apparent stability of the fear → surprise misclassification may be partially driven by paradigm selection rather than representing a fully generalizable phenomenon.

Data completeness also poses a challenge: three studies included only partial or truncated data, and two additional studies were unavailable. Efforts to recover these data sources are planned and may improve the robustness of future updates.

The search strategy introduces additional constraints. Screening of titles and abstracts was restricted to English and Portuguese, potentially excluding relevant studies published in other languages. Notably, the facial emotion recognition (FER) literature includes substantial contributions from Mandarin-, Japanese-, and Korean-speaking research groups (e.g., labs in Beijing and Kyoto), which may be underrepresented in the present pool. We estimate that this restriction may have excluded approximately 5–15 eligible studies.

Furthermore, the search was limited to PubMed/MEDLINE and Scopus; discipline-specific databases — notably PsycINFO, the reference index for psychology, and Web of Science, which indexes some high-

quality journals not covered elsewhere — as well as Embase and the Cochrane Library, were not searched. We relied on the broad, multidisciplinary indexing of Scopus, which overlaps substantially with these sources, to mitigate this gap, but cannot exclude that a small number of studies indexed only in PsycINFO or Web of Science were missed. No grey-literature search (thesis, dissertations, conference proceedings, or repository/preprint platforms) was conducted. To partially offset these constraints, we performed backward citation searching for the reference lists of all included studies (see below). These choices were made because the convergence of our principal findings with prior reviews nonetheless suggests that the resulting risk of selection bias is limited.

Methodologically, data extraction was performed by a single reviewer with targeted audits, but without formal dual independent coding. While auditing reduces error risk, some residual extraction inaccuracies may persist.

Regarding publication bias, multiple approaches were applied. Egger's test ($p = 0.196$), Begg's rank correlation (Kendall's $\tau = 0.24$, $p = 0.14$), and the Duval and Tweedie trim-and-fill method ($L_0 = 1$ imputed study) all converged on no detectable asymmetry. However, these tests have limited statistical power given the relatively small number of studies ($k = 21$). The near absence of studies reporting fear recognition accuracy below 20% may therefore still reflect underrepresentation that is not captured by formal bias detection methods.

At the domain level, risk of bias analysis revealed that Domain D2 (Stimuli) accumulated the highest number of HIGH-risk ratings (5 out of 40 studies), primarily due to incomplete reporting of stimulus standardization. This limitation is particularly relevant for interpreting cultural moderation effects, as heterogeneity in stimuli remains only partially measured.

Finally, no formal meta-regression analyses were conducted to jointly test potential moderators³⁹. Instead, subgroup analyses were used as informal moderation approaches. Future updates incorporating a larger number of studies could allow for more robust meta-regression modelling.

Future Directions

Five priorities emerge for future research^{9,43}. First, large-scale cross-cultural studies with identical paradigms^{35,36} that separate culture from method. Second, stimulus databases with openly available confusion matrices^{29,33,66}. Third, longitudinal studies tracking fear recognition across clinical stages to test the marker hypothesis^{9,24}. Fourth, RCTs of fear-surprise discrimination training in autism or PTSD⁶⁹. Fifth, future research on fear recognition in healthy and clinical adults should incorporate a gender perspective and an intersectional approach; to better understand how sex/gender and other sociodemographic and clinical factors may influence fear facial recognition performance and related confusion patterns .

CONCLUSION

This systematic review of 40 studies confirms three convergent findings of the literature¹²: (i) fear accuracy in healthy adults varies massively between 19% and 94%, with a global pool of 54.5% and $I^2 \approx 88\%$, a dispersion driven largely by methodological factors — chiefly the number of response options — rather than by true differences in ability^{29,32,42}; (ii) the fear → surprise confusion is statistically stable at $\approx 32\%$ with $I^2 \approx 4\%$ ^{6,19,29}; and (iii) fear remains the most difficult facial emotion to recognize^{1,25}. Taken together, these findings recommend: (a) replacing aggregate pools with subgroup analyses³⁹; (b) reporting the complete confusion matrix as a norm⁶; (c) using unbiased-hit-rate metrics rather than raw accuracy¹¹; and (d) testing in future studies whether the fear non-surprise profile is a sensitive clinical marker once the response-option set is held constant²⁶. In conclusion, fear recognition is a clinically sensitive ability whose fear–surprise confusion may serve as a differential marker; assessed with standardized methods and detailed error analyses, it could inform the diagnosis and follow-up of neuropsychiatric disorders.

METHODS

Protocol and reporting standards

This systematic review was conducted and reported in accordance with the PRISMA 2020 statement¹², with adaptations for meta-analysis of proportions following⁴⁷. The protocol for this review was registered in PROSPERO (CRD420261436532; <https://www.crd.york.ac.uk/PROSPERO/view/CRD420261436532>). Eligibility criteria, search strategy, data extraction template and analytic plan were defined *a priori*, before the start of data extraction and analysis. All working files (the triage spreadsheet, the data extraction template, the risk-of-bias workbook, and the meta-analysis scripts) are made available as supplementary material. This study was carried out within the scope of the broader AI.EVENTS project (ALGARVE-FEDER 01180500).

Eligibility criteria

Eligibility followed the pre-specified PICO(S) criteria summarized in Supplementary Table S1¹². Studies were eligible if they assessed facial recognition of the fear emotion in adults (≥ 18 years), reported a recognition metric for fear (raw accuracy, Hu unbiased hit rate, d' , reaction time, detection threshold, or neurophysiological correlate of fear-related processing), and used original empirical methods.

Studies were excluded if they met any of the following pre-defined criteria:

- Inadequate population (animal studies, or studies restricted to participants < 18 years).
- Absence of fear assessment (the fear emotion was not evaluated or was inseparably grouped with other emotions).
- Systematic reviews, meta-analyses, or secondary literature reviews (these were used as theoretical references but not as primary sources).

- Emotion recognition based exclusively on speech or text (for example, recognition through prosody without a facial component).
- Exclusive use of artificial intelligence or machine learning approaches without human behavioral assessment.
- Exclusively clinical samples without a healthy comparison group, unless the primary contrast was between clinical subgroups assessed with comparable methods.
- Duplicated studies.

Information sources and search strategy

The literature search was conducted in PubMed/MEDLINE and Scopus across two phases. The first phase, conducted by the coordinating author, combined Boolean terms for fear, facial expression, recognition and emotion, restricted to 2016–2025. The two databases returned 302 records (PubMed/MEDLINE) and 2,533 (Scopus), for 2,835 records in total; after removal of 126 duplicates, 2,709 records were screened at title and abstract, of which 46 passed to full-text retrieval. The second phase, conducted by the lead author, assessed full-text eligibility for these 46 records. PDFs were retrieved through institutional and open-access sources (37 records) and through the coordinating author (5 additional PDFs); the full text of 2 records could not be retrieved and a further 2 were only partially available (Fig. 1). No additional database searches were performed. Backward citation searching (screening of the reference lists of all included studies) was subsequently conducted; it identified no further studies meeting the eligibility criteria. The final search across all databases was conducted on 30 March 2026.

Study selection

Study selection proceeded in two stages. Title/abstract screening and full-text eligibility were each assessed independently by two reviewers (the coordinating and lead authors). Five borderline studies^{20,33,59,66,72} were ultimately included after consensus, and disagreements were resolved by discussion. Figure 1 shows the PRISMA flow diagram¹².

Data extraction

A structured data extraction template with 85 fields in nine sections (bibliographic identification; study design; population; stimuli; paradigm; FER metrics; neural correlates; moderators; limitations) was pilot tested on five calibration studies and validated with the coordinator before application to the remaining 35 studies. Data extraction was performed independently by two reviewers, with discrepancies resolved by consensus. Two targeted audits were additionally conducted: the first^{25,29} identified four denominator errors in²⁵, all corrected; the second^{22,24} revealed no material errors. Fields with NR, N/A or TBD values were explicitly coded as such and not imputed. Two studies without retrievable PDFs^{57,58} were recorded as placeholders but excluded from the quantitative analyses.

Risk of bias assessment

Risk of bias was assessed using a hybrid tool adapted from the Prediction Model Risk Of Bias Assessment Tool ⁴⁵ and the Appraisal Tool for Cross-Sectional Studies ⁴⁶. The hybrid design was motivated by the inadequacy of standard tools for this study type: PROBAST was originally designed for prediction-model studies, AXIS for cross-sectional studies, and QUADAS-2 for diagnostic accuracy. None was originally intended for psychophysical studies of emotion recognition. Our adaptation comprises 14 signaling questions distributed across four domains (Supplementary Table S2): Participants (4 items), Stimuli (4 items), Outcome (3 items), and Analysis (3 items). All 14 items, their decision anchors, and the scoring rules were defined as a priori, before risk-of-bias assessment began.

Note

Each item was coded as LOW, UNCLEAR, HIGH or N/A. Domain and overall classification followed the cumulative rule: HIGH if any item was HIGH; LOW only if every item was LOW; UNCLEAR otherwise. UNCLEAR items were resolved against these pre-specified disambiguation rules when the relevant information was implicit in the study design. Applying the rules raised the share of LOW item-level judgements to 61% (343 of 560 applicable items) while leaving every overall classification unchanged, showing that the categorical results are driven by the exigent cumulative rule (LOW only if every item is LOW) rather than by any post-hoc adjustment. To preserve informativeness, we computed a continuous Quality Score (number of LOW items / numbers of applicable items), ranging from 0 to 1. The assessment was conducted independently by two reviewers, with disagreements resolved by consensus.

Quantitative synthesis: meta-analysis of fear accuracy in healthy controls

Inclusion criteria for the pool

Pool inclusion required: (a) fear accuracy in healthy controls with a clear numerical value; (b) reported HC sample size; (c) a reported or imputable SD; and (d) a forced-choice recognition paradigm. Twenty-one studies satisfied all four criteria. Of the 19 excluded from the pool, 13 used incompatible paradigms (z-scores: ^{27,71}; Likert intensity ratings: ^{20,56,65,73,74}; QUEST thresholds: ⁵⁵; bodily sensation maps: ⁷²; EEG-only without behavioral accuracy: ^{34,59}; FACS production: ⁷⁵; auxiliary FER: ⁶⁴, 3 reported only partial numerical data ^{23,38,60}, and 2 had no retrievable PDF ^{57,58}. All 19 excluded studies are included in the narrative synthesis (Section 3.8) and listed in Supplementary Table S2.

Statistical model

The pool was computed using the random-effects model of DerSimonian and Laird (1986) ⁴⁸ with a logit transformation of the proportions, as recommended by Barendregt et al. (2013) ⁴⁷ for meta-analyses of prevalence and similar bounded proportions. The logit transform stabilizes the variance and avoids the

problem of proportions close to the boundaries (0 or 1) having an underestimated binomial variance. The central formulas are:

- Transformation : $\text{logit}(p) = \log(p / (1 - p))$.
- Variance of the proportion: $\text{var}(p) = p(1 - p)/n$ (binomial), or $\text{var}(p) = \text{SD}^2/n$ if a reported SD implies a larger standard error.
- Variance of the logit: $\text{var}(\text{logit}) = \text{var}(p) / [p(1 - p)]^2$.
- Between-study variance τ^2 : DerSimonian–Laird method-of-moments estimator, $\tau^2 = \max(0, (Q - \text{df}) / c)$, where $c = \sum w_i - \sum w_i^2 / \sum w_i$.
- Random-effects weights: $w_{i-RE} = 1 / (\text{var}(\text{logit})_i + \tau^2)$.
- Pooled estimate: $\mu_{-RE} = \sum (w_{i-RE} \cdot \text{logit}(p_i)) / \sum w_{i-RE}$; $\text{SE} = \sqrt{1 / \sum w_{i-RE}}$; 95% CI = $\text{expit}(\mu \pm 1.96 \cdot \text{SE}) \times 100$.
- Heterogeneity: $I^2 = \max(0, (Q - \text{df}) / Q)$; Cochran's $Q = \sum w_{i-FE} \cdot (\text{logit}(p_i) - \mu_{-FE})^2$. I^2 was interpreted following Higgins et al. (2003).

A random-effects rather than fixed-effects model was chosen because clinical and methodological heterogeneity were expected and were ultimately confirmed ($I^2 \approx 88\%$).

Imputation of missing standard deviations

For studies that reported fear accuracy without an explicit SD, the SD across subjects was imputed using the binomial approximation $\text{SD} = \sqrt{p \cdot (1 - p) / k}$, where p is the mean proportion correct, and k is the number of trials per participant. Equivalently, the standard error of the pool entry was taken as $\max(\text{SE}_{\text{binomial}}, \text{SE}_{\text{reported}})$, where $\text{SE}_{\text{binomial}} = \sqrt{p \cdot (1 - p) / n_{\text{HC}}}$ is the conservative binomial standard error for an aggregate proportion across all n_{HC} participants. This is a conservative approximation because it assumes independence within participants. An imputation flag was preserved for sensitivity analyses.

Subgroup analyses

We pre-specified five subgroup analyses to investigate moderators of fear accuracy:

Stimulus modality

static (photograph or single-apex morph) vs dynamic (video, neutral-to-apex morph, or multimodal).

Cultural background

Western (North American, European, South American, or Australian) vs East Asian (Chinese, Japanese, or Korean).

Era of publication

≤ 2019 (pre-COVID-19) vs ≥ 2020 (pandemic era, with potential prolonged exposure to facial masks).

Number of response options

< 6 vs ≥ 6 , testing specifically whether the inclusion of surprise as an option materially alters accuracy.

Inclusion of a neutral response category (post-hoc and reported descriptively only).

Each subgroup pool was computed using the same DerSimonian–Laird logit procedure. A formal test of moderation (e.g., meta-regression with subgroup interaction) was not conducted because of the small k in some subgroups; we instead rely on non-overlap of confidence intervals as an informal indicator of effect.

Analysis of cross-emotion confusions

To investigate with which other emotions fear is most frequently confused in healthy controls, we identified studies that published a complete confusion matrix (i.e., the proportion of responses for each target category given a fear stimulus). Three studies satisfied this criterion^{6,29,49}. For each pair $\text{fear} \rightarrow X$ (where $X \in \{\text{anger, disgust, happiness, neutral, sadness, surprise}\}$), a random-effects logit pool was computed using the same procedure described in Section 2.7. In parallel, a categorical synthesis was conducted across 22 studies that reported qualitative information on the dominant confusion direction (simple count of studies identifying each emotion as the primary confusion partner of fear).

Sensitivity analyses

Publication bias was assessed using a funnel plot and Egger's regression⁶³. Two further sensitivity analyses were performed: a leave-one-out analysis (sequentially excluding each of the 21 studies and re-estimating the pool) and a Quality-Score-restricted pool (restricted to studies with QS $\geq$ pool median). Both are reported in Section 3.7 and Supplementary Fig. S3.

Software and reproducibility

All analyses were conducted in Python 3.10 with the library's pandas (data manipulation), NumPy (numerical operations), SciPy. Stats (regression and statistical tests), matplotlib (figures), and openpyxl (Excel I/O). Scripts for extraction tracking, risk-of-bias scoring, meta-analysis, and figure generation are available in the project repository. The Excel files MetaAnalysis_Input.xlsx, MetaAnalysis_Computed.xlsx, Subgroups.xlsx, Confusions_FearTo_Quantitative.xlsx, and Confusions_Pooled.xlsx are provided to permit direct replication in any meta-analytic software (e.g., R metafor, Stata metan).

Ethics statement

This research is a systematic review and meta-analysis of previously published studies; no new primary data from human or animal participants were collected. All included studies were conducted by their respective authors with institutional ethical approval, as reported in the original articles. Because the present synthesis uses only previously published, de-identified data, no additional ethical approval was required. The review was conducted in accordance with the principles of the Declaration of Helsinki (World Medical Association) and the PRISMA 2020 statement ¹².

Declarations

Declaration of interest

The authors report there are no competing interests to declare.

Funding

This work was supported by FCT.IP and by the AI.EVENT project (ALGARVE-FEDER 01180500, Ref. 17325), co-financed by ALGARVE 2030, Portugal 2030, and the European Union.

Data availability statement

The data, code and supplementary materials supporting the findings of this study have been deposited in the Open Science Framework. The repository is currently private and will be made publicly available on acceptance with a permanent DOI. Reviewers and the Editor may request access by contacting the corresponding author at frcarvalho@ualg.pt.

Author contributions (CRediT)

A. T. Martins and **A. Gomes** contributed to study conception, methodology and supervision, and defined eligibility criteria and validated borderline inclusion decisions. **J. A. Camacho-Ruiz, J. Russo, and C. Campos** conducted study selection at full text. **F. R. Carvalho** performed data extraction, risk-of-bias assessment, statistical analysis and meta-analysis. **C. M. Galvez-Sánchez contributed** to manuscript revision. J. Rodrigues contributed funding acquisition, resources and supervision. All authors critically revised the manuscript and approved the final version. All authors participated in the research, the analysis, and the writing of the manuscript, in both the original draft and the revised version. All authors meet the ICMJE criteria for authorship.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used an AI assistant to support manuscript editing. After using it, the authors reviewed and edited the content and take full responsibility for the content of the publication.

Included studies 2016- 2025 (n = 40) citation and title

Author	Article title
Abutalebi et al., 2024 ⁴²	On the brain struggles to recognize basic facial emotions with face masks: an fMRI study
Aho-Ozhan et al., 2016 ²¹	Perception of Emotional Facial Expressions in Amyotrophic Lateral Sclerosis (ALS) at Behavioural and Brain Metabolic Level
Bell et al., 2017 ²⁸	Earthquake Brain: Altered Recognition and Misclassification of Facial Expressions Are Related to Trauma Exposure but Not Posttraumatic Stress Disorder
Bologna et al., 2016 ²³	Altered Kinematics of Facial Emotion Expression and Emotion Recognition Deficits Are Unrelated in Parkinson's Disease
Burgund, 2021 ⁷⁶	Left Hemisphere Dominance for Negative Facial Expressions: The Influence of Task
Calvo et al., 2018 ⁶	Selective eye fixations on diagnostic face regions of dynamic emotional expressions: KDEF-dyn database
Chen et al., 2025 ²⁶	Dynamic facial expression recognition deficits in schizophrenia: A classification study
Chong et al., 2024 ³³	A Cantonese Audio-Visual Emotional Speech (CAVES) dataset
Chung et al., 2019 ⁶¹	Development and Validation of the Yonsei Face Database (YFace DB)
Colonnello et al., 2019 ⁶⁴	First Impression Misleads Emotion Recognition
Dodich et al., 2022 ²²	Emotion Recognition Deficits in the Differential Diagnosis of Amnestic Mild Cognitive Impairment: A Cognitive Marker for the Limbic-Predominant Phenotype
Esposito et al., 2016 ⁶²	A pilot study on the decoding of dynamic emotional expressions in major depressive disorder
Franca et al., 2023 ³⁷	Seeing emotions in the eyes: a validated test to study individual differences in the perception of basic emotions
Gu et al., 2023 ³¹	Eye movements and ERP biomarkers for face processing problems in avoidant attachment-style individuals
Hama & Koeda, 2023 ³²	Characteristics of healthy Japanese young adults with respect to recognition of facial expressions: a preliminary study
Iannaccone et al., 2025 ²⁴	Exploring the role of epileptic focus lateralization on facial emotion recognition in the spectrum of mesial temporal lobe epilepsy
Kim H et al., 2023 ⁷⁵	Facial Expressions Track Depressive Symptoms in Old Age

Author	Article title
Kim M et al., 2022 ⁶⁶	Effects of diagnostic regions on facial emotion recognition: The moving window technique
Kosonogov & Titova, 2019 ⁴⁹	Recognition of all basic emotions varies in accuracy and reaction time: A new verbal method of measurement
Lannoy et al., 2019 ³⁸	Disrupted Fear and Sadness Recognition in Binge Drinking: A Combined Group and Individual Analysis
Lee SC et al., 2022 ⁷⁷	Error patterns of facial emotion recognition in patients with schizophrenia
Lomoriello et al., 2023 ⁶⁵	Sensitivity to Basic Emotional Expressions and the Emotion Perception Space in the Absence of Facial Mimicry: The Case of Individuals With Congenital Facial Palsy
Lyons et al., 2021 ⁷²	Bodily Maps of Emotion in Major Depressive Disorder
Manno et al., 2025 ⁵⁸	Observer and stimulus factors jointly shape perceptual similarity of static and dynamic facial emotions
Meléndez et al., 2020 ⁶⁰	Emotion recognition changes in a confinement situation due to COVID-19
Mo et al., 2021 ²⁵	Confusion Effects of Facial Expression Recognition in Patients with Major Depressive Disorder and Healthy Controls
Puffet & Rigoulot, 2025 ²⁹	Validation of the Emotionally Congruent and Incongruent Face-Body Static Set (ECIFBSS)
Richer et al., 2025 ⁵⁶	Is It Pain, Anger, Disgust, or Sadness? Individual Differences in Expectations of Pain Facial Expressions
Rodger et al., 2018 ⁵⁵	Quantifying facial expression signal and intensity use during development
Salvato et al., 2023 ²⁷	"I feel your fear": superior fear recognition in organized crime members
Sarma & Barma, 2021 ⁵⁹	Usefulness of blinking duration variability (BDV) in discriminating emotional states
Schönenberg et al., 2016 ⁷⁴	Facial Affect Recognition in Violent and Nonviolent Antisocial Behavior Subtypes
Schönenberg et al., 2019 ⁷³	Processing of Dynamic Affective Information in Adults With ADHD
Schwartz et al., 2018 ²⁰	Unaltered emotional experience in Parkinson's disease: Pupillometry and behavioral evidence

Author	Article title
Suslow et al., 2023 ¹¹	Interpersonal problems and recognition of facial emotions in healthy individuals
Szmulewicz et al., 2020 ⁷¹	Emotional processing subtypes in bipolar disorder: A cluster analysis
Tabesh et al., 2025 ⁵⁷	Confidently recognizing facial emotion: the impact of masks and attachment styles
Uono et al., 2017 ⁷⁸	Putamen Volume is Negatively Correlated with the Ability to Recognize Fearful Facial Expressions
Xie et al., 2023 ³⁴	Brain Activation Differences of Six Basic Emotions Between 2D Screen and Virtual Reality Modalities
Yamasaki & Miyata, 2025 ⁴¹	Effects of Interoceptive Awareness on Recognition of and Sensitivity to Emotions in Masked Facial Stimuli

References

1. Russell, J. A. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychol. Bull.* **115**, 102–141 (1994).
2. The Expression of Emotion in Man and Animals, by & Darwin, C.
https://www.gutenberg.org/files/1227/1227-h/1227-h.htm#link2H_4_0003
3. Ekman, P. & Friesen, W. V. Constants across cultures in the face and emotion. *J. Pers. Soc. Psychol.* **17**, 124–129 (1971).
4. Adolphs, R., Tranel, D., Damasio, H. & Damasio, A. Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature* **372**, 669–672 (1994).
5. Ekman, P. & Cordaro, D. What is meant by calling emotions basic. *Emot. Rev.* **3**, 364–370 (2011).
6. Calvo, M. G., Fernández-Martín, A., Gutiérrez-García, A. & Lundqvist, D. Selective eye fixations on diagnostic face regions of dynamic emotional expressions: KDEF-dyn database. *Sci Rep* **8**, (2018).
7. Susskind, J. M. et al. Expressing fear enhances sensory acquisition. *Nat. Neurosci.* **11**, 843–850 (2008).
8. Bediou, B. et al. Impaired social cognition in mild Alzheimer disease. *J. Geriatr. Psychiatry Neurol.* **22**, 130–140 (2009).
9. Argaud, S., Vérin, M., Sauleau, P. & Grandjean, D. Facial emotion recognition in Parkinson's disease: A review and new hypotheses. *Mov. Disord.* **33**, 554–567 (2018).
10. Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M. & Moberg, P. J. Facial emotion perception in schizophrenia: a meta-analytic review. *Schizophr Bull.* **36**, 1009–1019 (2010).
11. Suslow, T., Lemster, A., Koelkebeck, K. & Kersting, A. Interpersonal problems and recognition of facial emotions in healthy individuals. *Front Psychiatry* **14**, (2023).

12. Page, M. J. et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International J. Surgery* **88**, (2021).
13. Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R. & Schyns, P. G. Facial expressions of emotion are not culturally universal. *Proc. Natl. Acad. Sci. U. S. A.* 109, 7241–7244 (2012).
14. Gendron, M., Roberson, D., van der Vyver, J. M. & Barrett, L. F. Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion* **14**, 251–262 (2014).
15. Cordaro, D. T. et al. Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion* **18**, 75–93 (2018).
16. Cowen, A. S. et al. Sixteen facial expressions occur in similar contexts worldwide. *Nature* 2020 589:7841 589, 251–257 (2020).
17. van Heijst, K., Kret, M. E. & Ploeger, A. Basic Emotions or Constructed Emotions: Insights From Taking an Evolutionary Perspective. *Perspect. Psychol. Sci.* **20**, 377–391 (2025).
18. Jack, R. E., Sun, W., Delis, I., Garrod, O. G. B. & Schyns, P. G. Four not six: Revealing culturally common facial expressions of emotion. *J. Exp. Psychol. Gen.* **145**, 708–730 (2016).
19. Calder, A. J. et al. Facial Emotion Recognition after Bilateral Amygdala Damage: Differentially Severe Impairment of Fear. *Cogn. Neuropsychol.* **13**, 699–745 (1996).
20. Schwartz, R., Rothermich, K., Kotz, S. A. & Pell, M. D. Unaltered emotional experience in Parkinson's disease: Pupillometry and behavioral evidence. *J. Clin. Exp. Neuropsychol.* **40**, 303–316 (2018).
21. Aho-Ozhan, H. E. A. et al. Perception of Emotional Facial Expressions in Amyotrophic Lateral Sclerosis (ALS) at Behavioural and Brain Metabolic Level. *PLoS One* **11**, (2016).
22. Dodich, A. et al. Emotion Recognition Deficits in the Differential Diagnosis of Amnesic Mild Cognitive Impairment: A Cognitive Marker for the Limbic-Predominant Phenotype. *J. Int. Neuropsychol. Soc.* **28**, 203–209 (2022).
23. Bologna, M. et al. Altered Kinematics of Facial Emotion Expression and Emotion Recognition Deficits Are Unrelated in Parkinson's Disease. *Front Neurol* **7**, (2016).
24. Iannaccone, F. et al. Exploring the role of epileptic focus lateralization on facial emotion recognition in the spectrum of mesial temporal lobe epilepsy. *Front Syst. Neurosci* **18**, (2025).
25. Mo, F., Gu, J., Zhao, K. & Fu, X. Confusion Effects of Facial Expression Recognition in Patients With Major Depressive Disorder and Healthy Controls. *Front Psychol* **12**, (2021).
26. Chen, N. et al. Dynamic facial expression recognition deficits in schizophrenia: A classification study. *Asian J. Psychiatr* **114**, (2025).
27. Salvato, G. et al. I feel your fear: superior fear recognition in organised crime members. *Cogn. Emot.* **37**, 430–438 (2023).
28. Bell, C. J. et al. Earthquake Brain: Altered Recognition and Misclassification of Facial Expressions Are Related to Trauma Exposure but Not Posttraumatic Stress Disorder. *Front Psychiatry* **8**, (2017).

29. Puffet, A. S. & Rigoulot, S. Validation of the Emotionally Congruent and Incongruent Face-Body Static Set (ECIFBSS). *Behav Res. Methods* **57**, (2025).
30. Adolphs, R. et al. A mechanism for impaired fear recognition after amygdala damage. *Nature* **433**, 68–72 (2005).
31. Gu, S. et al. Eye movements and ERP biomarkers for face processing problems in avoidant attachment-style individuals. *Front Behav. Neurosci* **17**, (2023).
32. Hama, T. & Koeda, M. Characteristics of healthy Japanese young adults with respect to recognition of facial expressions: a preliminary study. *BMC Psychol* **11**, (2023).
33. Chong, C. S., Davis, C. & Kim, J. A. Cantonese Audio-Visual Emotional Speech (CAVES) dataset. *Behav. Res. Methods*. **56**, 5264–5278 (2024).
34. Xie, J., Lan, P., Wang, S., Luo, Y. & Liu, G. Brain Activation Differences of Six Basic Emotions Between 2D Screen and Virtual Reality Modalities. *IEEE Trans. Neural Syst. Rehabil. Eng.* **31**, 700–709 (2023).
35. Elfenbein, H. A. & Ambady, N. On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychol. Bull.* **128**, 203–235 (2002).
36. Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R. & Schyns, P. G. Facial expressions of emotion are not culturally universal. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 7241–7244 (2012).
37. Franca, M., Bolognini, N. & Brysbaert, M. Seeing emotions in the eyes: a validated test to study individual differences in the perception of basic emotions. *Cogn. Res. Princ Implic.* **8**, 67 (2023).
38. Lannoy, S. et al. Disrupted Fear and Sadness Recognition in Binge Drinking: A Combined Group and Individual Analysis. *Alcohol Clin. Exp. Res.* **43**, 1978–1985 (2019).
39. Higgins, J. P. T., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. *BMJ* **327**, 557–560 (2003).
40. Carbon, C. C. Wearing Face Masks Strongly Confuses Counterparts in Reading Emotions. *Front. Psychol.* **11**, 566886 (2020).
41. Yamasaki, K. & Miyata, H. Effects of Interoceptive Awareness on Recognition of and Sensitivity to Emotions in Masked Facial Stimuli. *Behavioral Sci. (Basel Switzerland)* **15**, (2025).
42. Abutalebi, J. et al. On the brain struggles to recognize basic facial emotions with face masks: an fMRI study. *Front Psychol* **15**, (2024).
43. Pavlova, M. A. & Sokolov, A. A. Reading Covered Faces. *Cereb. Cortex.* **32**, 249–265 (2022).
44. Calvo, M. G., Averó, P., Fernández-Martín, A. & Recio, G. Recognition thresholds for static and dynamic emotional faces. *Emotion* **16**, 1186–1200 (2016).
45. Wolff, R. F. et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. (2019). <https://doi.org/10.7326/M18-1376> **170**, 51–58
46. Downes, M. J., Brennan, M. L., Williams, H. C. & Dean, R. S. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ Open.* **6**, e011458 (2016).
47. Barendregt, J. J., Doi, S. A., Lee, Y. Y., Norman, R. E. & Vos, T. Meta-analysis of prevalence. *J. Epidemiol. Community Health.* **67**, 974–978 (2013).

48. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control Clin. Trials*. **7**, 177–188 (1986).
49. Kosonogov, V. & Titova, A. Recognition of all basic emotions varies in accuracy and reaction time: A new verbal method of measurement. *Int. J. Psychol.* **54**, 582–588 (2019).
50. Malykhin, N. et al. Emotional recognition across the adult lifespan: Effects of age, sex, cognitive empathy, alexithymia traits, and amygdala subnuclei volumes. *J. Neurosci. Res.* **101**, 367–383 (2023).
51. Griffiths, S. et al. Impaired Recognition of Basic Emotions from Facial Expressions in Young People with Autism Spectrum Disorder: Assessing the Importance of Expression Intensity. *J. Autism Dev. Disord.* **49**, 2768–2778 (2019).
52. Martino, D. J., Samamé, C. & Strejilevich, S. A. Stability of facial emotion recognition performance in bipolar disorder. *Psychiatry Res.* **243**, 182–184 (2016).
53. Dalkiran, M. et al. The Change in Facial Emotion Recognition Ability in Inpatients with Treatment Resistant Schizophrenia After Electroconvulsive Therapy. *Psychiatr. Q.* **88**, 535–543 (2017).
54. Jin, Y. et al. Altered emotional recognition and expression in patients with Parkinson's disease. *Neuropsychiatr Dis. Treat.* **13**, 2891–2902 (2017).
55. Rodger, H., Lao, J. & Caldara, R. Quantifying facial expression signal and intensity use during development. *J. Exp. Child. Psychol.* **174**, 41–59 (2018).
56. Richer, A., Gingras, F., Plouffe-Demers, M. P., Fiset, D. & Blais, C. Is It Pain, Anger, Disgust, or Sadness? Individual Differences in Expectations of Pain Facial Expressions. *Emotion* (2025). <https://doi.org/10.1037/emo0001516> doi:10.1037/emo0001516.
57. Tabesh, S., Zia-Tohidi, A., Firoozi, M. & Maghami-Sharif, Z. Confidently recognizing facial emotion: the impact of masks and attachment styles. *Current Psychology* 2025 44:18 44, 15215–15226 (2025).
58. Manno, L. et al. Observer and stimulus factors jointly shape perceptual similarity of static and dynamic facial emotions. *Cogn. Emot.* <https://doi.org/10.1080/02699931.2025.2578697> (2025).
59. Sarma, P. & Barma, S. Usefulness of blinking duration variability (BDV) in discriminating emotional states. *Biomed. Signal. Process. Control.* **69**, 102883 (2021).
60. Meléndez, J. C. et al. Emotion recognition changes in a confinement situation due to COVID-19. *J Environ. Psychol* **72**, (2020).
61. Chung, K. M., Kim, S., Jung, W. H. & Kim, Y. Development and Validation of the Yonsei Face Database (YFace DB). *Front Psychol* **10**, (2019).
62. Esposito, A., Scibelli, F. & Vinciarelli, A. A pilot study on the decoding of dynamic emotional expressions in major depressive disorder. in *Smart Innovation, Systems and Technologies* vol. 54 189–200Springer Science and Business Media Deutschland GmbH, (2016).
63. Egger, M., Smith, G. D., Schneider, M. & Minder, C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* **315**, 629–634 (1997).

64. Colonnello, V., Russo, P. M. & Mattarozzi, K. First Impression Misleads Emotion Recognition. *Front Psychol* **10**, (2019).
65. Lomoriello, A. S. et al. Sensitivity to Basic Emotional Expressions and the Emotion Perception Space in the Absence of Facial Mimicry: The Case of Individuals With Congenital Facial Palsy. (2023). <https://doi.org/10.1037/emo0001275.supp> doi:10.1037/emo0001275.supp.
66. Kim, M., Cho, Y. & Kim, S. Y. Effects of diagnostic regions on facial emotion recognition: The moving window technique. *Front Psychol* **13**, (2022).
67. Ekman, P. & Friesen, W. V. Facial Action Coding System. *PsycTESTS Dataset*. <https://doi.org/10.1037/T27734-000> (2019).
68. Jack, R. E., Garrod, O. G. B. & Schyns, P. G. Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Curr. Biol.* **24**, 187–192 (2014).
69. Schönenberg, M. et al. Addressing perceptual insensitivity to facial affect in violent offenders: first evidence for the efficacy of a novel implicit training approach. *Psychol. Med.* **44**, 1043–1052 (2014).
70. Adolphs, R., Tranel, D., Damasio, H. & Damasio, A. R. Fear and the human amygdala. *J. Neurosci.* **15**, 5879–5891 (1995).
71. Szmulewicz, A., Millett, C. E., Shanahan, M., Gunning, F. M. & Burdick, K. E. Emotional processing subtypes in bipolar disorder: A cluster analysis. *J. Affect. Disord.* **266**, 194–200 (2020).
72. Lyons, N. et al. Bodily Maps of Emotion in Major Depressive Disorder. *Cogn. Therapy Res.* **2020**. **45:3** **45**, 508–516 (2020).
73. Schönenberg, M., Schneidt, A., Wiedemann, E. & Jusyte, A. Processing of Dynamic Affective Information in Adults With ADHD. *J. Atten. Disord.* **23**, 32–39 (2019).
74. Schönenberg, M., Mayer, V., Christian, S., Louis, S. & Jusyte, A. K. *Brief Report FACiAL AffeCt ReCoGNitioN iN VioLeNt AND NoNVioLeNt ANtiSoCiAL BeHAVior SUBtYpeS*. *Journal Personality Disorders* **29** (2015).
75. Kim, H. et al. Facial Expressions Track Depressive Symptoms in Old Age. *Sens.* **2023**. **23**, 7080 (2023).
76. Burgund, E. D. Left Hemisphere Dominance for Negative Facial Expressions: The Influence of Task. *Front Hum. Neurosci* **15**, (2021).
77. Lee, S. C. et al. Error patterns of facial emotion recognition in patients with schizophrenia. *J. Affect. Disord.* **300**, 441–448 (2022).
78. Uono, S. et al. Putamen Volume is Negatively Correlated with the Ability to Recognize Fearful Facial Expressions. *Brain Topogr.* **30**, 774–784 (2017).

Figures

Figure 1. PRISMA 2020 flow diagram — Fear Recognition Systematic Review

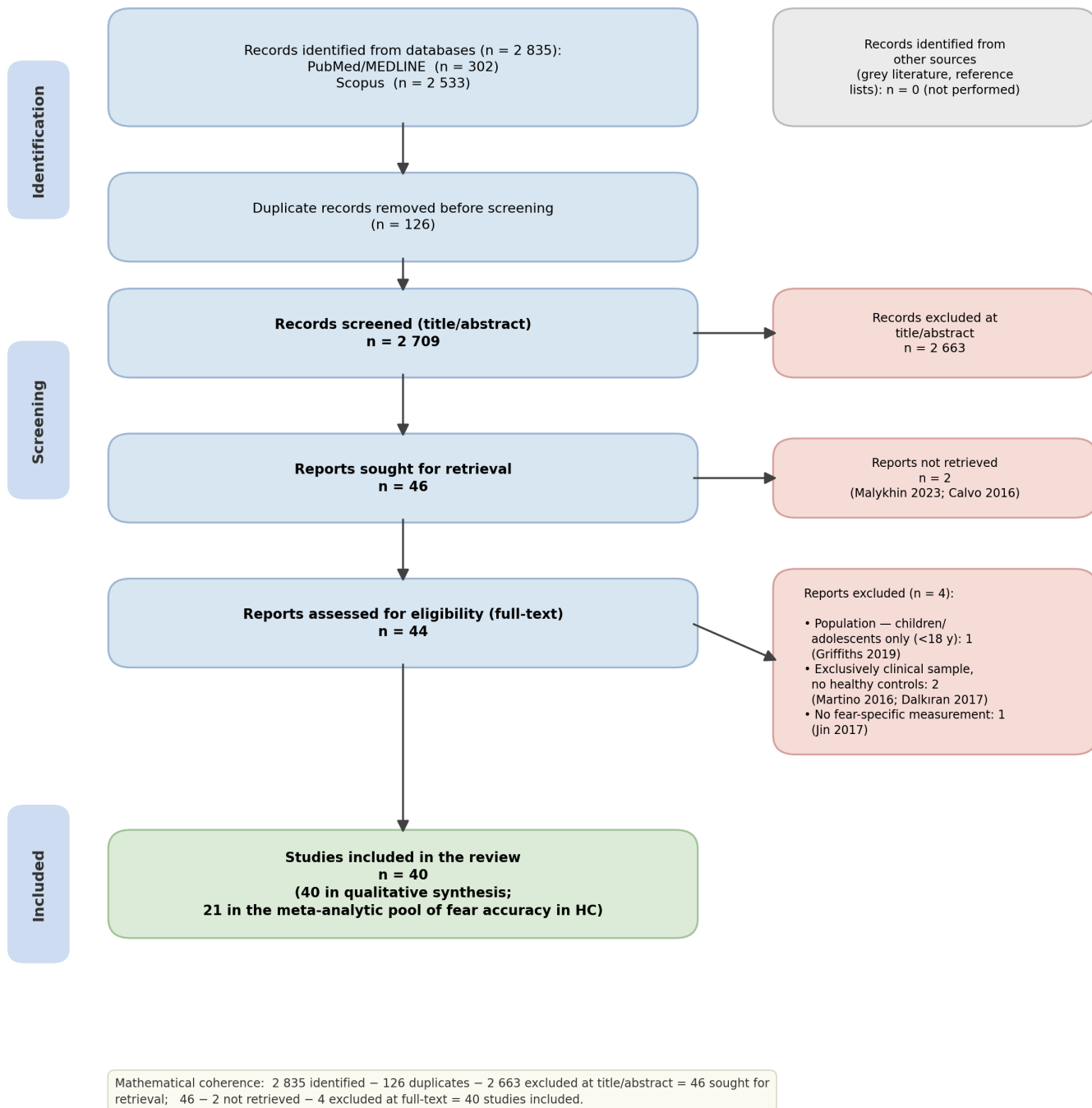


Figure 1

PRISMA 2020 flow diagram (Page et al., 2021). 2,835 records were identified (302 PubMed/MEDLINE; 2,533 Scopus); after removal of 126 duplicates, 2,709 were screened and 2,663 excluded at title/abstract, leaving 46 sought for retrieval. Of these, 2 were not retrieved and 4 were excluded at full-text screening (with reasons), yielding 40 included studies.

Fig.3 — Forest plot: HC fear accuracy (k=21 studies, n=1369 HC)
Pool: 54.5% [95% CI 45.8-62.9], I²=88%, τ^2 =0.53

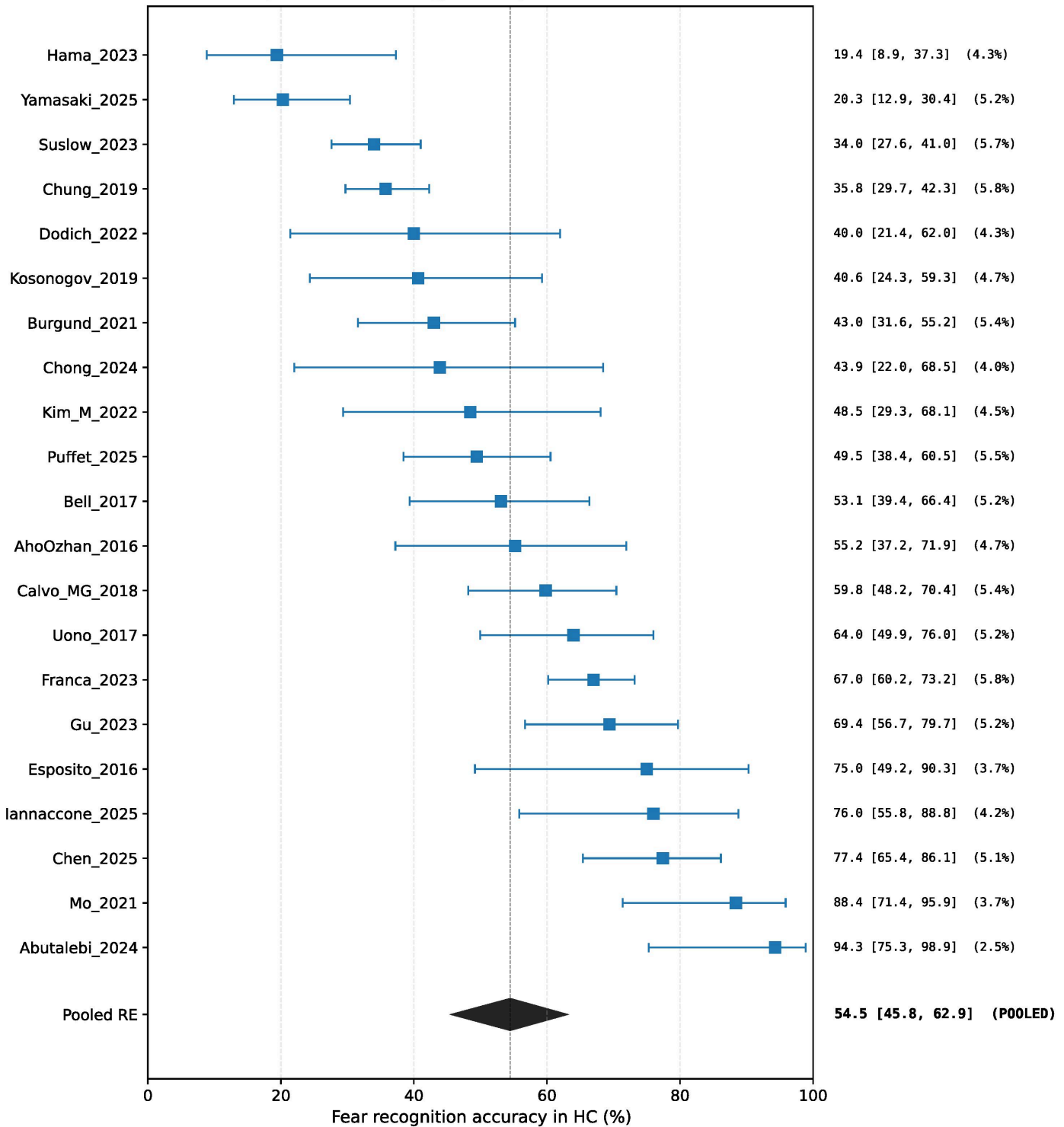


Figure 2

Forest plot of fear recognition accuracy in healthy controls (k = 21; total n = 1369). Pool RE = 54.5% [45.8–62.9]; I² = 88%; τ^2 = 0.53 (DerSimonian & Laird, 1986; Higgins et al., 2003).

Fig.4 — Subgroup analyses (k=21 studies, n=1369 HC)

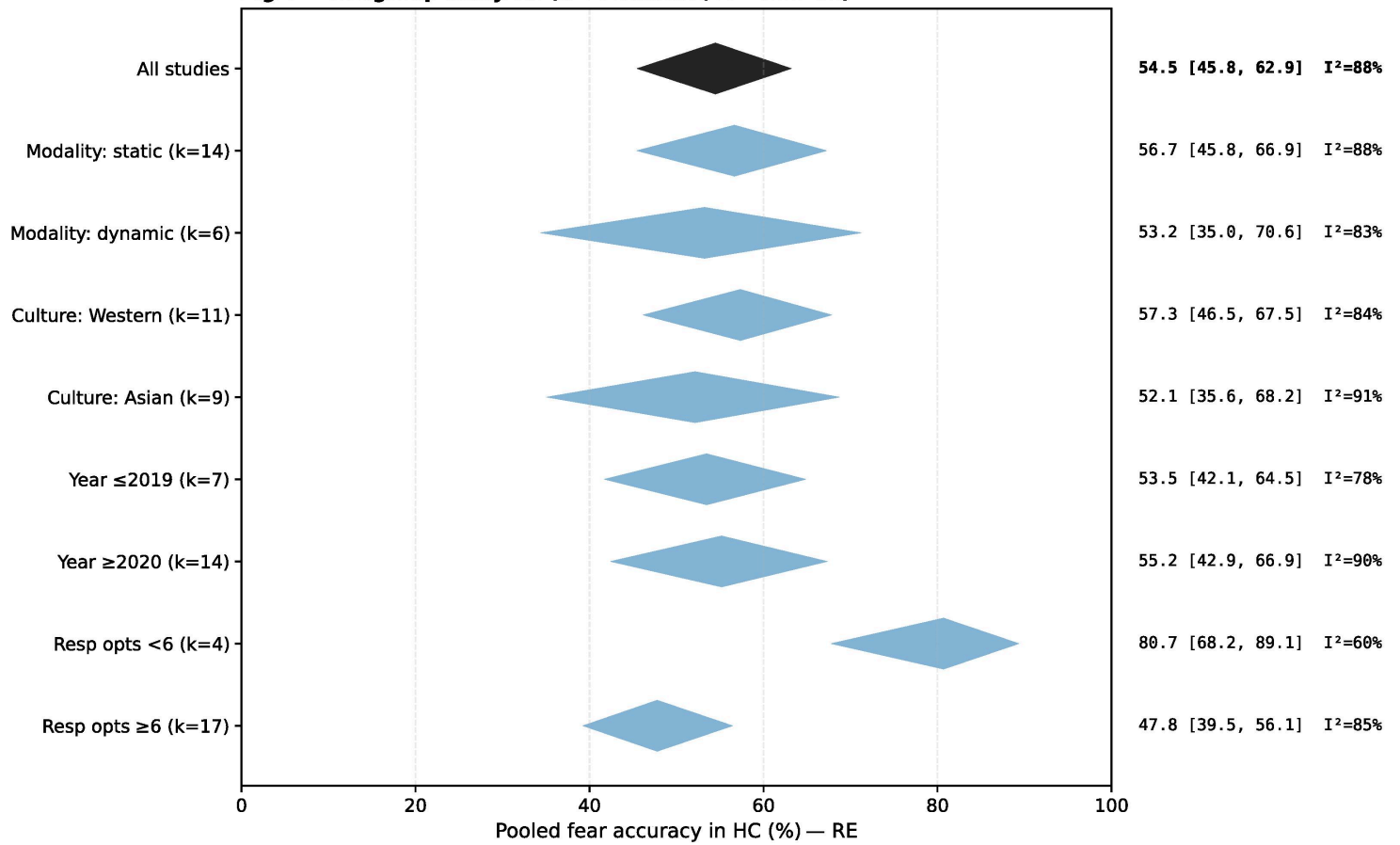


Figure 3

Subgroup analyses. Largest moderator: response options < 6 (80.7%) vs ≥ 6 (47.8%).

Fig.6 — Cross-emotion confusion patterns for fear in healthy controls
fear → surprise is the dominant confusion (perceptual-equivalence hypothesis)

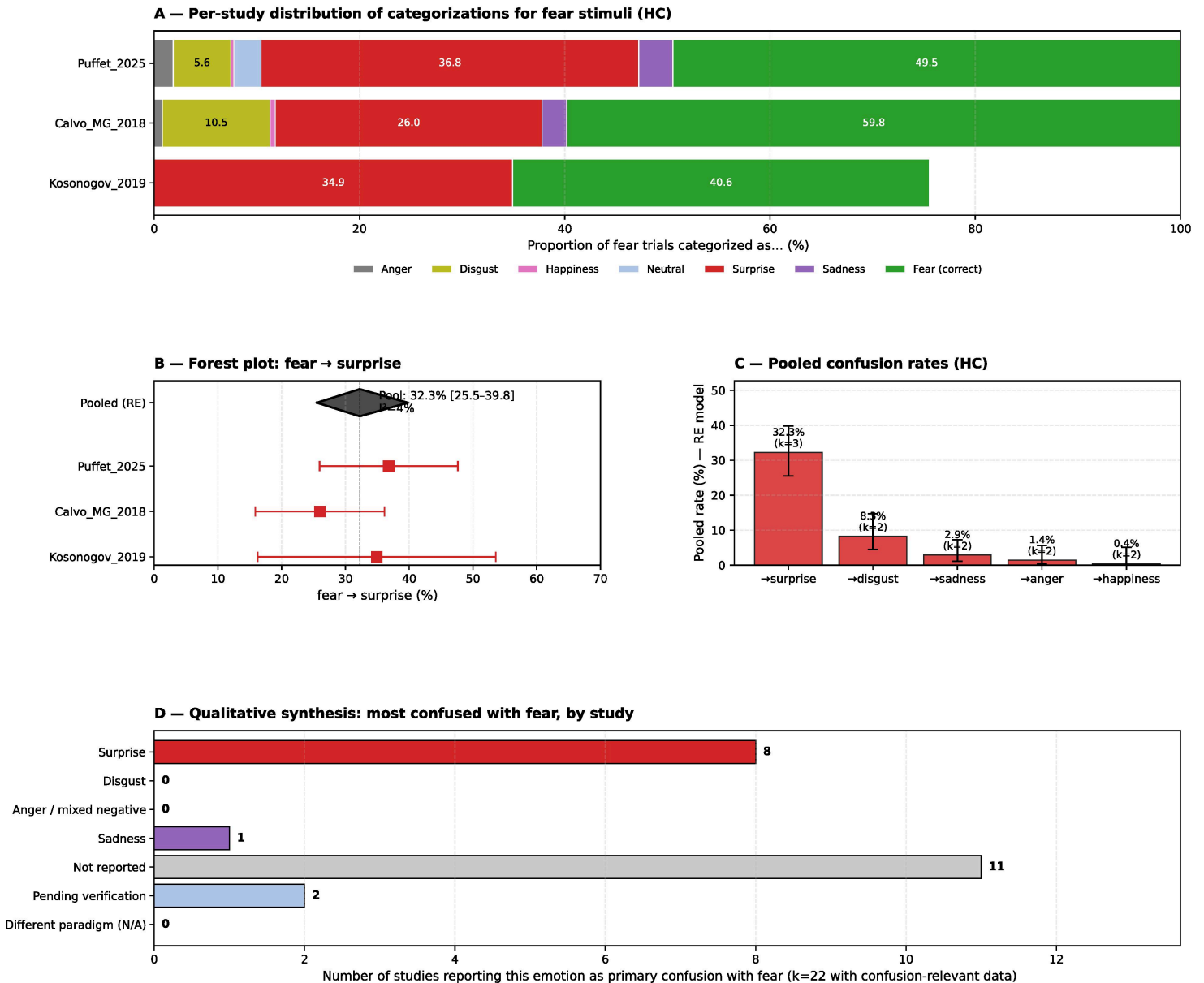


Figure 4

Cross-emotion confusion patterns for fear in healthy controls (Puffet et al., 2025; Calvo et al., 2018; Kosonogov & Titova, 2019). Panel A: per-study distribution of categorisations for fear stimuli; Panel B: forest plot of the fear → surprise pool; Panel C: pooled rates across all confusion targets; Panel D: qualitative synthesis (most-confused-with-fear, by study).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformation.docx](#)