Supplemental Material S1. Summary of the included studies in the scoping review.

Author (year)	Country	Study aims	Study type	Intervention study?	Task paradigm (stimulus type)	ASD sample size	ASD sample demographics (age: years)	Presence of comparison group?	Quality index score	Key findings[a]
Macdonald et al. (1989)	The UK	To assess the recognition and expression of emotional cues in both facial and vocal modalities	Behavioral study	No	Affect naming (V, A)	10	Adult males (*M* = 27.2, *SD* = 5.6)	Yes	0.91	The socio-emotional deficit in autistic subjects is evident on tests of both expression and recognition, and is not modality specific.
Boucher et al. (2000)	The UK	To assess the possible role of a cross-modal matching impairment as a cause, or contributory cause, of a voice-face affect matching impairment	Behavioral study	No	vocal affect naming, vocal-facial affect matching (V, A)	19	Children (*M* = 9.58, *SD* = 1.0)	Yes	0.86	Children with autism are impaired relative to language-matched typically developing children on the test of affect matching. Children with moderate to high-functioning autism do not have cross-modal processing impairments.

Author (year)	Country	Study aims	Study type	Intervention study?	Task paradigm (stimulus type)	ASD sample size	ASD sample demographics (age: years)	Presence of comparison group?	Quality index score	Key findings[a]
O’Connor et al. (2007)	New Zealand	To examine the ability of adults with AS and age-matched TD controls to identify incongruent and congruent emotional information from the face and voice	Behavioral study	No	Cross-modal (in)congruent emotion identification and discrimination (V, A, V+A)	18	Adults (*M* = 26.9, *SD* = 7.8)	Yes	0.77	Adults with AS are less accurate at discriminating incongruent from congruent expressive faces and voices relative to TD subjects.
Kahana-Kalman et al. (2008)	The US	To examine emotion recognition abilities of young children with ASD	Behavioral study	No	intermodal matching (V+A)	18	Young children (*M* = 4.08)	Yes	0.82	Emotion recognition is not systematically deficient in children with autism. There is no difference in intermodal (visual and auditory) matching of maternal emotional expressions between children with autism and normally developing children.

Author (year)	**Country**	**Study aims**	**Study type**	**Intervention study?**	**Task paradigm (stimulus type)**	**ASD sample size**	**ASD sample demographics (age: years)**	**Presence of comparison group?**	**Quality index score**	**Key findings**[a]
Philip et al. (2010)	The UK	To investigate whether individuals with ASD have pervasive deficits in emotion processing across stimulus domains	Behavioral study	No	Forced choice identification (V, A)	23	Adults (*M* = 32.5, *SD* = 10.9)	Yes	0.91	There are significant and broad-ranging deficits in emotion processing in ASD present across a range of stimulus domains and in the auditory and visual modality.
Jones et al. (2011)	The UK	To test both visual (facial) and auditory (verbal and non-verbal vocalizations) emotion recognition in adolescents with ASD compared to age- and IQ-matched controls	Behavioral study	No	Forced choice identification (V, A)	99	Adolescents (*M* = 15.5, *SD* = 5.6 months)	Yes	0.91	No evidence of a fundamental emotion recognition deficit has been found in the ASD group. Basic emotion recognition ability should not be considered in isolation as the source of the social and communication difficulties observed in ASD.

Author (year)	**Country**	**Study aims**	**Study type**	**Intervention study?**	**Task paradigm (stimulus type)**	**ASD sample size**	**ASD sample demographics (age: years)**	**Presence of comparison group?**	**Quality index score**	**Key findings[a]**
Magnée et al. (2011)	The Netherlands	To investigate how manipulation of attention affected the integration of visual and auditory emotional information	Electrophysiological study	No	Cross-modal (in)congruent emotion processing with attention manipulation (V, A, V+A)	23	Adult males (*M* = 22.7, *SD* = 3.8)	Yes	0.82	The multisensory processing of emotional signals in ASD is intact under appropriate circumstances. Atypical multisensory processing in ASD is shown to be secondary to attentional manipulation.

Author (year)	**Country**	**Study aims**	**Study type**	**Intervention study?**	**Task paradigm (stimulus type)**	**ASD sample size**	**ASD sample demographics (age: years)**	**Presence of comparison group?**	**Quality index score**	**Key findings[a]**
Vannetzel et al. (2011)	France	To explore processing of neutral and emotional human stimuli (by auditory, visual and multimodal channels) in children with PDD-NOS compared to TD children	Behavioral study	No	Emotion discrimination (V, A, V+A)	10	Children (*M* = 9.6, *SD* = 1.7)	Yes	0.91	Children with PDD-NOS present global emotional human stimuli processing difficulties, which dramatically contrast with their ability to process neutral human stimuli. They have difficulties comprehending emotion and partially compensate for this problem using multimodal processing.

Author (year)	**Country**	**Study aims**	**Study type**	**Intervention study?**	**Task paradigm (stimulus type)**	**ASD sample size**	**ASD sample demographics (age: years)**	**Presence of comparison group?**	**Quality index score**	**Key findings**[a]
Lopata et al. (2012)	The US	To evaluate the feasibility and initial efficacy of a manualized comprehensive school-based intervention	Behavioral study	Yes	Forced choice identification (V, A)	12	Elementary school children (*M* = 7.33, *SD* = 0.98)	No	0.62	There are significant increases in the children's ability to identify emotional states in facial and vocal expressions after the comprehensive school-based intervention.
Stewart et al. (2012)	The UK	To examine the connection between vocal and facial recognition of emotion and to test whether a semantic compensatory strategy could be observed in emotion detection in speech stimuli only	Behavioral study	No	Lexical-prosodic (in)congruent emotion identification (V, A)	11	Adults (*M* = 27.2, *SD* = 7.5)	Yes	0.82	In decoding emotion from spoken utterances, individuals with ASC rely more on verbal semantics than TD individuals, presumably as a strategy to compensate for their difficulties in using prosodic cues to recognize emotions.

Author (year)	**Country**	**Study aims**	**Study type**	**Intervention study?**	**Task paradigm (stimulus type)**	**ASD sample size**	**ASD sample demographics (age: years)**	**Presence of comparison group?**	**Quality index score**	**Key findings[a]**
Charbonneau et al. (2013)	Canada	To explore the perception and the integration of emotion expressions in ASD	Behavioral study	No	Forced two-choice discrimination (V, A, V+A)	32	Adolescents and adults (*M* = 21, *SD* = 6)	Yes	0.95	There is an altered sensitivity to emotion expressions in ASD population that is not modality-specific. Autistic participants benefit from exposure to bimodal information to a lesser extent than did the TD group, indicative of a decreased multisensory gain in this population. There are joint alterations for both the perception and the integration of multisensory emotion expressions in ASD.

Author (year)	Country	Study aims	Study type	Intervention study?	Task paradigm (stimulus type)	ASD sample size	ASD sample demographics (age: years)	Presence of comparison group?	Quality index score	Key findings[a]
Doi et al. (2013)	Japan	To investigate the ability of adults with AS to recognize emotional categories of facial expressions and emotional prosody with graded emotional intensities and to clarify the underlying cause of the deficits in emotion recognition ability in adults with AS	Behavioral study	No	Four-choice identification (V, A)	20	Adult males with AS (*M* = 32.1, *SD* = 7.3)	Yes	0.86	The difference between the AS group and the TD group in emotion recognition from facial expression and from prosodic information might derive at least partly from modality-specific processing of low-level perceptual features.
Kandalaft et al. (2013)	The US	To investigate the feasibility of a 10-session Virtual Reality Social Cognition Training intervention in adults with HFA	Behavioral study	Yes	Forced choice identification (V, A)	8	HFA adults (*M* = 21.25, *SD* = 2.71)	No	0.73	Significant increases on social cognitive measures of theory of mind and emotion recognition, as well as in real life social and occupational functioning were found post-training.

Author (year)	**Country**	**Study aims**	**Study type**	**Intervention study?**	**Task paradigm (stimulus type)**	**ASD sample size**	**ASD sample demographics (age: years)**	**Presence of comparison group?**	**Quality index score**	**Key findings[a]**
Lerner et al. (2013)	The US	To elucidate heterogeneity in emotion processing and to assess the presence of multimodal deficits in emotion perception among youth with ASD	Behavioral and electrophysiological study	No	Rating & four-choice identification (V, A)	34	School-aged children and adolescents (*M* = 13.07, *SD* = 2.07)	No	0.86	Many youths with ASD do possess multimodal deficits in emotion recognition. The essential multimodality of emotion recognition in individuals with ASD may derive from early social information processing speed, despite heterogeneous behavioral performance.

Author (year)	Country	Study aims	Study type	Intervention study?	Task paradigm (stimulus type)	ASD sample size	ASD sample demographics (age: years)	Presence of comparison group?	Quality index score	Key findings[a]
Matsuda et al. (2013)	Japan	To examine whether young children with ASD could be taught to comprehend the relationship between affective prosody and visually presented facial expressions via cross-modal matching-to-sample training	Behavioral study	Yes	Vocal-facial affect matching (V, A)	4	Young children and school-aged children (*M* = 5.5)	No	0.54	Cross-modal matching-to-sample training procedures can be suitable for teaching cross-modal emotion perception skills to younger children with ASD.
Singh et al. (2014)	The US	To investigate sensitivity to prosodic and semantic cues to emotion in individuals with HFA	Behavioral study	No	Lexical-prosodic (in)congruent emotion identification (A)	10	Children (*M* = 10.58)	Yes	0.86	Participants with HFA are impaired in the spontaneous integration of prosodic and semantic cues to emotion. Insensitivity to surface detail, such as prosody, in HFA appears to be highly task dependent and selective to the domain of emotion.

Author (year)	**Country**	**Study aims**	**Study type**	**Intervention study?**	**Task paradigm (stimulus type)**	**ASD sample size**	**ASD sample demographics (age: years)**	**Presence of comparison group?**	**Quality index score**	**Key findings[a]**
Segal et al. (2014)	Israel	To assess how adolescents with autism who vary in the severity of autistic characteristics judge the emotional state of the speaker when lexical and prosodic information is congruent or incongruent	Behavioral study	No	Lexical-prosodic (in)congruent emotion identification (A)	24	Adolescents (*M* = 15.03)	Yes	0.82	Adolescents with ASD were able to accurately perceive the emotions of the speaker based on lexical or prosodic information alone. The severity of autistic characteristics influenced the ability to give more weight to the prosodic over the lexical information.

Author (year)	Country	Study aims	Study type	Intervention study?	Task paradigm (stimulus type)	ASD sample size	ASD sample demographics (age: years)	Presence of comparison group?	Quality index score	Key findings[a]
Globerson et al. (2015)	Israel	To examine the relative contribution of psychoacoustic factors and more general emotion recognition abilities to affective prosody recognition in ASD	Behavioral study	No	Forced choice identification (V, A)	20	Males (*M* = 28.8, *SD* = 6.8)	Yes	0.86	There are multimodal emotion recognition deficits in ASD. Alongside general, cross-modal, emotion recognition abilities, auditory perceptual abilities play a significant and potentially compensatory role in prosody recognition in ASD.
Golan et al. (2015)	Israel	To compare emotion recognition abilities of children with ASC and typically developing controls and to examine the psychometric properties of the CAM-C battery	Behavioral study	No	Four-choice identification (V, A)	30	Children (*M* = 9.7, *SD* = 1.2)	Yes	0.91	8- to 11-year-old children with ASC have difficulties in complex emotion and mental state recognition in both faces and voices.

Author (year)	**Country**	**Study aims**	**Study type**	**Intervention study?**	**Task paradigm (stimulus type)**	**ASD sample size**	**ASD sample demographics (age: years)**	**Presence of comparison group?**	**Quality index score**	**Key findings[a]**
Taylor et al. (2015)	Australia	To investigate facial and vocal emotion recognition in children with ASD, children with specific language impairment and TD children	Behavioral study	No	Four-choice identification (V, A)	29	Children (*M* = 8.86)	Yes	0.86	Individuals with ASD are impaired in facial and vocal affect recognition. There are differences in emotion recognition abilities between ASD children with normal language and those with impaired language.

Author (year)	**Country**	**Study aims**	**Study type**	**Intervention study?**	**Task paradigm (stimulus type)**	**ASD sample size**	**ASD sample demographics (age: years)**	**Presence of comparison group?**	**Quality index score**	**Key findings**[a]
Xavier et al. (2015)	France	To explore unimodal and multimodal emotion processing in children with ASD	Eye-tracking study	No	Forced choice identification (V, A, V+A)	19	Children (*M* = 9.95, *SD* = 1.75)	Yes	0.86	Multisensory processing allowed children with ASD to partially compensate for the difficulties that were experienced in the visual modality. Developmental age was significantly associated only with the multimodal task for children with ASD. Language impairments tended to be associated with emotion recognition scores of ASD children in the auditory modality.

Author (year)	Country	Study aims	Study type	Intervention study?	Task paradigm (stimulus type)	ASD sample size	ASD sample demographics (age: years)	Presence of comparison group?	Quality index score	Key findings[a]
Golan et al. (2018)	Israel	To assess the relative contribution of cues from several perceptual modalities to facial emotion recognition in children with ASD	Behavioral study	No	Cross-modal emotion matching (V+A)	29	Children (*M* = 9.13, *SD* = 1.18)	Yes	0.86	Overall facial emotion recognition deficits in ASD. Children with ASD struggled with face-face matching, compared to voice-face and word-face combinations. Cross-modal integration is preferable to intra-modal processing in emotion recognition. Performance of ASD in the voice-face cross-modal recognition task was related to adaptive communication skills.
Su et al. (2018)	China	To explore deficits in multimodal emotion recognition and eye gaze in children with ASD	Eye-tracking study	No	Affect naming (V+A)	10	Children (*M* = 8.85, *SD* = 1.35)	Yes	0.77	Children with ASD have deficits in multimodal emotion recognition.

Author (year)	Country	Study aims	Study type	Intervention study?	Task paradigm (stimulus type)	ASD sample size	ASD sample demographics (age: years)	Presence of comparison group?	Quality index score	Key findings[a]
Scheerer et al. (2020)	Canada	To investigate whether autistic children have difficulty extracting affect from prosody, and whether this difficulty might be related to social competence	Behavioral study	No	Cross-modal emotion matching (V+A)	26	Children (*M* = 10.02, *SD* = 1.79)	Yes	0.91	Children with ASD can accurately extract the affective meaning conveyed by changes in prosody but were less accurate at matching the voice-clips to the emotional faces, suggesting that autistic children struggle to make use of this information in a social context.

Notes. V = visual stimuli; A = auditory stimuli; V+A = visual and auditory stimuli presented simultaneously; AS = Asperger’s syndrome; HFA = high-functioning autism; PDD-NOS = pervasive developmental disorder not otherwise specified; TD = typically developing; CAM-C = Cambridge Mindreading Face-Voice Battery for Children.

aKey findings relating to the multi-channel processing of emotion in ASD.