CORTEX 154 (2022) 322–332

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research Report

Egocentric biases are predicted by the precision of self-related predictions

Leora Sevi [a,*], *Mirta Stantic* [a], *Jennifer Murphy* [b], *Michel-Pierre Coll* [c], *Caroline Catmur* [d] *and Geoffrey Bird* [a,e,**]

[a] *Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom*
[b] *Department of Psychology, Royal Holloway, University of London, Egham, Surrey, United Kingdom*
[c] *School of Psychology, Laval University, Quebec City, Canada*
[d] *Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom*
[e] *Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom*

ARTICLE INFO

Article history:
Received 27 June 2021
Reviewed 28 September 2021
Revised 16 January 2022
Accepted 26 April 2022
Action editor Michael Schwartze
Published online 9 June 2022

Keywords:
Emotion recognition
Predictive coding
Empathy
Generative model
Precision
Predictive interoceptive coding

ABSTRACT

According to predictive processing theories, emotional inference involves simultaneously minimising discrepancies between predictions and sensory evidence relating to both one's own and others' states, achievable by altering either one's own state (empathy) or perception of another's state (egocentric bias) so they are more congruent. We tested a key hypothesis of these accounts, that predictions are weighted in inference according to their precision (inverse variance). If correct, increasingly precise self-related predictions should be associated with increasingly biased perception of another's emotional expression. We manipulated predictions about upcoming own-pain (low or high magnitude) using cues that afforded either precise (a narrow range of possible magnitudes) or imprecise (a wide range) predictions. Participants judged pained facial expressions presented concurrently with own-pain to be more intense when own-pain was greater, and precise cues increased this biasing effect. Implications of conceptualising interpersonal influence in terms of predictive processing are discussed.

* *Corresponding author*. Department of Experimental Psychology, University of Oxford, Anna Watts Building, Woodstock Road, Oxford, OX2 6GG, United Kingdom.
** *Corresponding author*. Department of Experimental Psychology, University of Oxford, Anna Watts Building, Woodstock Road, Oxford, OX2 6GG, United Kingdom.
E-mail addresses: leora.sevi@psy.ox.ac.uk (L. Sevi), geoff.bird@psy.ox.ac.uk (G. Bird).
https://doi.org/10.1016/j.cortex.2022.04.021

1. Introduction

The notion that the brain is an inferential machine that actively models the state of the world to estimate the causes of (i.e., explain) the evidence it encounters is becoming increasingly influential in cognitive neuroscience. Within the predictive processing framework (Clark, 2016; Friston, 2010; Hohwy, 2013), the brain continuously tests the predictions generated by its models against incoming sensory evidence, and discrepancies between predictions and sensory evidence (prediction errors) are used to update the models such that they better fit the sensory evidence. Within a particular moment, this process of updating predictions such that they best match the current state of the world corresponds to perception. The influence of prior predictions on what is perceived is a function of their expected uncertainty, or reliability, relative to prediction errors. This expected uncertainty is termed precision, and is a measure of signal-to-noise or confidence (mathematically, the inverse variance of a signal). Higher precision of prior predictions relative to prediction errors will result in a percept that is biased away from sensory evidence towards the prior to a greater extent than will lower relative precision, which will result in a percept that is more consistent with sensory evidence.

Since the brain must model the state not only of the external world but also of the body, predictive processing theories have been extended to include interoceptive (e.g., hunger, satiety, pain) predictions as well as those in the exteroceptive and proprioceptive domains (Allen, Levy, Parr, & Friston, 2019; Barrett & Simmons, 2015; Seth, 2013; Seth & Friston, 2016). The hierarchical structure of generative models allows these different modalities to be integrated; prediction errors from contingent events in each modality are assimilated at higher levels such that they contextualise one another. Higher levels can therefore represent more abstract hidden causes of sensory evidence (Linson, Parr, & Friston, 2020; Pezzulo, Rigoli, & Friston, 2015) such as our action goals and intentions. Under the assumption that these hidden states cause the same sensory consequences in others as they do ourselves, deep generative models are thought to permit the prediction and inference of the action goals (Kilner, Friston, & Frith, 2007), mental states (Friston & Frith, 2015; Koster-Hale & Saxe, 2013) and affective states (Demekas, Parr, & Friston, 2020; Ondobaka, Kilner, & Friston, 2017; Peng, Huang, Liu, & Cui, 2019; Quattrocki & Friston, 2014) of other people.

If there are contingencies between states of the self and (perceivable cues indicating) the states of others, our internal models will encode these contingencies, linking interoceptive/proprioceptive predictions regarding the state of the self, with exteroceptive predictions indicating the state of the other. Assuming that most people are empathic, the typical developmental environment will be one in which others (e.g., caregivers) respond to one's own states by displaying that same state themselves (e.g., a caregiver will make a pained face when an infant is in pain). This means that the contingencies captured by internal models will be more likely to link states of the self with cues indicating the same state in others (Bird & Viding, 2014; Heyes & Bird, 2007).

Prediction error is reduced across multiple levels (and modalities) simultaneously, meaning that predictions regarding the state of the self can affect the perception of the state of another, and vice versa. For example, when experiencing pain, my internal model, attempting to estimate the causes of the interoceptive evidence I am receiving, will generate exteroceptive predictions regarding the state you will appear to be in (i.e., I will predict that you will display a pained expression). If you are not empathic and do not display any pain, then exteroceptive prediction errors would be generated, which could be resolved by biasing my perception of your expression such that it appears more pained, a form of 'emotional projection', or egocentric bias. Meanwhile, upon seeing my pain, your model, if it is the same as mine, will be generating interoceptive predictions regarding what your pain will feel like, which, when unfulfilled, may bias your perception of your own state.

If predictive processing theories are correct, then the extent to which perception of another's expression is biased towards one's own state should vary as a function of the precision of predictions regarding one's own state. Although several studies demonstrate that one's own state can influence perception of another's state (Edey, Yon, Cook, Dumontheil, & Press, 2017; Pezzulo et al., 2018; Rütgen et al., 2015, 2021; Silani, Lamm, Ruff, & Singer, 2013) and vice versa (Chapon, Perchet, Garcia-Larrea, & Frot, 2019; Liu et al., 2019), and turn-taking in songbirds has been successfully modelled using the predictive processing framework (Friston & Frith, 2015), empirical evidence for the role of precision in these interpersonal effects remains scarce. Effects of contextualised perception in the case of others' emotions have been successfully modelled (Mirza, Cullen, Parr, Shergill, & Moran, 2021), but not when the contextualising prior is one's own state. The present study was therefore designed to test the prediction that increased precision of predictions about one's own pain will be associated with increasingly biased perception of another's pain.

Participants were asked to judge the intensity of pained and happy facial expressions which were presented visually at the same time as pain was delivered. The magnitude of upcoming pain was signalled to participants using cues which indicated either a narrow or wide range of possible intensities, corresponding to high or low precision. In other words, participants' expectations about the pain they were about to receive, understood as probability distributions, had either a smaller or greater variance. Under the Bayesian models of perception described above, the precision of interoceptive predictions regarding pain should determine the effect of the prior on perception of one's own pain – more precise priors will bias pain perception to a greater degree (see Hoskin et al., 2019 for an empirical demonstration). Crucially, increasing precision of pain expectation should also be associated with increasing precision of exteroceptive predictions concerning the expression of pain in the other, and therefore a greater influence of those exteroceptive predictions on perception of the other's pain. Accordingly, it was hypothesised that the receipt of painful electrical stimulation would bias perception of another's pain state to a greater extent when accompanied by precise pain predictions, than when accompanied by imprecise pain predictions.

2. Methods

We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

2.1. Participants

Due to a lack of available data from which to calculate effect sizes, it was not possible to conduct power calculations. An opportunity sample was therefore collected, whereby all participants fulfilling the inclusion criteria who responded to the advertisement over six months of data collection were tested. Participants were excluded if they reported uncorrected impaired vision, a diagnosis of a neurodevelopmental disorder, meeting the criterion for severe alexithymia, taking any prescription medications with stimulant, sedative, or analgesic effects, or if they did not rate the maximum electrical stimulation as at least an 8 out of 10 (details below). From the participants tested, four were excluded from analyses, two because their performance on the measure of pain rating consistency was more than three standard deviations below the group mean, and two because their level of habituation was more than three standard deviations greater than the group mean (see below for details of these measures). Exclusion criteria were established prior to data analysis. The final sample was composed of 25 females and 24 males between the ages of 18 and 43 years ($M = 23.5$, $SD = 5.86$).

All participants gave written informed consent, and the study was approved by the Central University Research Ethics Committee, University of Oxford. Participants received a small honorarium for their participation.

2.2. Electrical stimulation and thresholding

Pain stimuli consisted of 200 μs electrical pulses generated by a Digitimer DS7A Constant Current Stimulator (Digitimer Ltd, Hertfordshire, United Kingdom). Stimuli were controlled by a custom MATLAB script and administered via a bar electrode (two disc electrodes with 9 mm diameter and 30 mm spacing) attached to the underside of the forearm of the non-dominant hand.

Stimulation levels were calibrated for each participant, creating a personalized '1' to '10' scale of pain. A value of '1' corresponded to a minimally painful pin-prick sensation, while '10' was the most painful stimulation participants were willing to receive up to 30 times over the following hour, which did not cause wincing, blinking, or a lapse in focus. Each participant received an ascending series of electrical stimulations, starting at an imperceptible level (1 mA), until they reported first feeling a painful pin-prick sensation. Starting from above this value, a series of stimulations of descending intensity was given until participants reported no longer feeling the pin-prick sensation. The ascending and descending painful thresholds were averaged to give the participant's '1' value. The intensity level was then further increased until the participant reported reaching '10'. Again, starting from above this value, a descending series of stimulations was given until participants reported the intensity dropping below '10', and the '10' value was taken as the average of the ascending and descending thresholds. The mean difference between '10' and '1' stimulation intensities was 40.1 mA ($SD = 22.0$). Provisional stimulation levels for values '2' through '9' were calculated as equidistant points between the '1' and '10' values. For each value, the provisional stimulation level was adjusted via further calibration according to participant feedback in increasingly fine intervals until the participant's subjective rating matched the assigned value.

2.3. Measures

2.3.1. Alexithymia

All participants completed the 20-item Toronto Alexithymia Scale (TAS-20; Bagby, Parker, & Taylor, 1994) so that participants meeting the criterion for severe alexithymia (TAS-20 score >60) could be excluded, as alexithymia has been associated with impaired interoception (Brewer, Cook, & Bird, 2016).

2.3.2. Pain reporting and degree of habituation

Before the main task, in a pre-test phase, participants received each of their 10 individually-calibrated stimulation intensities twice. The order of intensities was random, but held constant across all participants so that any effects of order on pain perception would be equal across participants. Participants were asked to rate each stimulation out of 10, based on the scale used during calibration. From these data, estimates of participant accuracy (correlation between the average of the two pre-test ratings and the actual intensity level) and consistency (correlation between the first and second pre-test rating for each stimulation level) were calculated. After the main task, in a post-test phase, this procedure was repeated, with each stimulation level being presented only once. Comparison of the pre- and post-test data allowed a measure of habituation to be derived (the mean difference between the post-test and the average of the pre-test ratings across intensity levels) for each participant.

2.4. Emotional facial expression stimuli

Stimuli were images of a female actor displaying happy and pained facial expressions of varying intensities, created by morphing each expression with a neutral expression using Morpheus Photo Morpher (Morpheus Development, Howell, Michigan). Original stimuli were obtained and validated by Simon, Craig, Gosselin, Belin, and Rainville (2008). Morphed images were converted to grey-scale and cropped into an oval shape to occlude hair, neck and peripheral information.

For both pain and happiness, 18 intermediate images between the neutral (0%) and the emotional expression (100%) were initially produced in 5% increments. A pilot study ($n = 50$) conducted using these images revealed that participants required 10% more happiness in happy morphs than the amount of pain required in pain morphs to judge the face as happy/pained, respectively. Therefore, to equalize perceived intensity of the two emotions, the final happy stimuli consisted of five morphs selected from a range of intensities (minimum 35%, maximum 70% intensity) each of which were

10% more intense than the corresponding pained morphs (minimum 25%, maximum 60% intensity; Fig. 1). Stimuli were 222 × 293 pixels in size, presented on a grey background in Psychtoolbox (Brainard, 1997) and viewed from a distance of approximately 60 cm. Presentation time was 425 msec.

2.5. *Pain cues*

In order to manipulate the precision of pain predictions, participants were presented with a cue prior to receiving each stimulation that informed them, with high or low precision, whether they were going to receive a high- or a low-pain stimulation. Cues were shown as horizontal bars, signifying the range from minimum (1) to maximum (10) pain, with a shaded region indicating the range of possible intensities of the upcoming stimulation. For low precision cues, this shaded region occupied 50% of the bar, indicating that the pain could be anywhere from minimum to mid-way (for low pain) or mid-way to maximum (for high pain). For high precision cues, 10% of the bar was shaded, centred around 25% (for low pain) and 75% (for high pain).

2.6. *Design*

The design consisted of three variables manipulated on a within-subjects basis: pain stimulation magnitude (Own-Pain: high or low), precision of pain expectation (Precision: high or low) and the type of expressed emotion (Emotion: pain or happiness) and trials representing the factorial combination of these three factors were presented equally over 120 trials in blocks of 24 trials. Blocks consisted of equal numbers of trials from every combination of experimental factors, presented in a random order. In low precision conditions, each face stimulus was paired once each with a stimulation of level '1', '3', and '5' (in the low own-pain condition) or a stimulation of level '6', '8' and '10' (in the high own-pain condition). In the high precision conditions, the stimulation given was always '3' in the low own-pain condition and '8' in the high own-pain condition. This ensured that the mean stimulation intensity received was equal across high and low precision conditions (i.e., '3' or '8') and also within each face stimulus.

2.7. *Procedure*

Participants were asked to have a full night's sleep before the experimental session, and to refrain from caffeine consumption on the day of testing. After obtaining informed consent, participants first completed the TAS-20. The electrode was then attached, the calibration procedure carried out, and the pre-test stimulation ratings obtained. There were six practice trials for the main task, presented in a random order but the conditions of which were fixed to include: each combination of Precision and Own-Pain conditions; the most extreme painful stimulations for low precision conditions (i.e., 1 and 5 for low own-pain and 6 and 10 for high own-pain), to reinforce the idea that low precision cues signal a wide range of potential upcoming pain relative to high precision cues, and the most and least intense face stimuli, so that participants could be instructed to calibrate their scale for rating the emotions accordingly (i.e., the least and most happy/pained expressions should correspond to '1' and '10' on the scale, respectively).

The structure of each trial of the main task is shown in Fig. 2. Participants were presented with the own-pain cue for three seconds before being presented with the facial expression for 425 msec. The electrical stimulation was delivered simultaneously with the presentation of the facial expression. Participants were then asked to judge the intensity of the emotional expression, the intensity of their own pain (both on

Fig. 1 – Experimental stimuli.*Note.* Pained (upper panel) and happy (lower panel) expressions, arranged by increasing intensity from left to right.

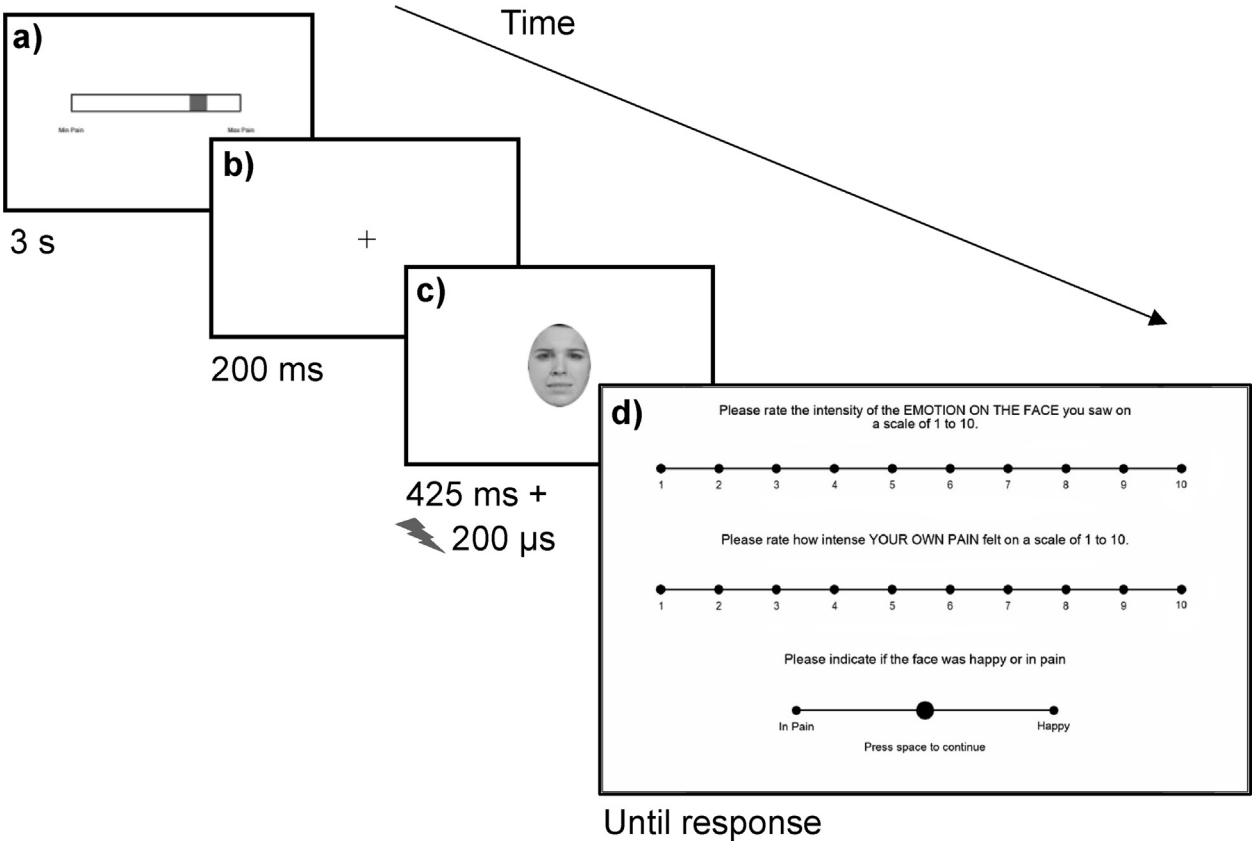

Fig. 2 – Task Structure. ***Note.*** **Example cue and stimulus shown – these varied across trials as specified under 'Design'. a) Cue: indicates the magnitude of the upcoming electrical stimulation (High or Low Own-pain) with either a High or Low degree of precision (High Pain, High Precision cue shown). b) ISI. c) Expression stimulus: either Pained or Happy with concurrent electrical stimulation. d) Response screen: Own-pain rating (1–10) + expression intensity rating (1–10) + emotion judgement (pained or happy).**

a scale of 1–10), and whether the facial expression was happy or pained. Participants were encouraged to take a break between blocks. After the main task, the post-test rating procedure was carried out. Study procedures were not formally pre-registered prior to the research being conducted.

3. Results

All statistical analyses were computed in JASP (Jasp Team, Amsterdam, The Netherlands). Analyses were not pre-registered prior to data collection. All tests are two-tailed unless otherwise specified. Bayesian analyses use JASP default priors. Participants' mean TAS-20 score was 41.8 ($SD = 9.04$).

3.1. *Pre- and post-test own-pain ratings*

The mean consistency correlation for own-pain rating (within-participant correlation between the two pre-test ratings) was .87 ($SD = .09$) and the mean accuracy correlation (within-participant correlation between the mean pre-test ratings and the calibrated pain levels) was .95 ($SD = .03$). The mean habituation score was .09 ($SD = .64$), corresponding to a very slight habituation.

3.2. *Expression intensity ratings*

One participant was an outlier in terms of the effect of own-pain on their ratings of happy faces in the imprecise condition (>3 *SD*s from the mean). Analyses excluding the outlier were performed in addition to those including it, which are reported below, with no differences in the patterns of significance.

Expression intensity ratings (see Fig. 3) were analysed using a 2 (Own-Pain: high *vs* low) × 2 (Precision: high *vs* low) × 2 (Emotion: pain *vs* happiness) repeated measures analysis of variance (ANOVA). As predicted, there was a significant 2-way interaction between Own-Pain and Emotion [$F(1, 48) = 5.61$, $p = .022$, $\eta_p^2 = .11$], and crucially, a significant 3-way interaction between Own-Pain, Precision and Emotion [$F(1, 48) = 11.4$, $p = .001$, $\eta_p^2 = .19$]. There were also significant main effects of Own-Pain [$F(1, 48) = 42.6$, $p < .001$, $\eta_p^2 = .47$] and Emotion [$F(1, 48) = 43.0$, $p < .001$, $\eta_p^2 = .47$]. All other main effects and interactions were non-significant and not of theoretical relevance [all $F(1, 48) \leq .47$, $p \geq .497$, $\eta_p^2 \leq .01$].

To deconstruct the 3-way interaction, two separate 2 × 2 repeated measures ANOVAs were performed for pain and happiness. To investigate these 2-way interactions and the significant 2-way interaction between Own-Pain and Emotion,

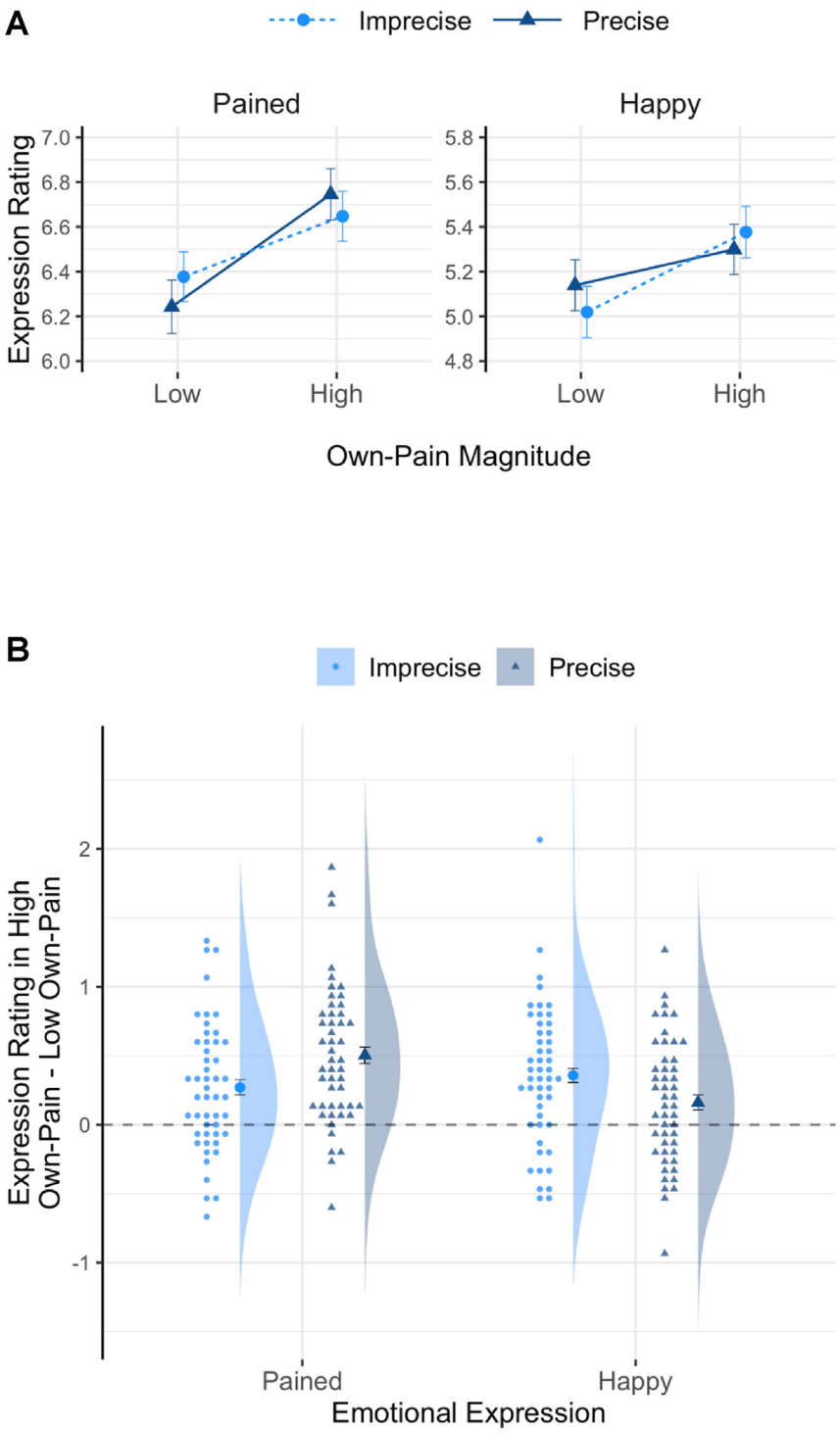

Fig. 3 – Expression Ratings as a Function of Own-Pain and Precision. *Note*. Panel A: mean rating of expression intensity as a function of own pain magnitude and precision for pained and happy expressions. Panel B: difference in expression rating between High and Low Own-Pain conditions for each combination of precision and emotion. Raincloud plots provide data distributions, mean values and raw data (jittered on the x axis). Error bars show within-subject standard error (Morey, 2008).

paired samples t-tests were performed and supplemented by Bayes factors (BF_{10}), using the framework proposed by Jeffreys (1961, see also Rouder, Speckman, Sun, Morey, & Iverson, 2009). The Bayes factors reflect how many times more likely the data are under the alternative hypothesis (that there is a difference in expression ratings between the relevant conditions) relative to the null (that there is no difference in expression ratings between the relevant conditions).

The simple main effect of Own-Pain on expression ratings ('mean difference' refers to expression ratings in high Own-

Pain subtracted from low Own-Pain conditions) was significant for both Emotion conditions, both across and within Precision conditions, but was greater for pained expressions [mean difference = .39, *SD* = .38; $t(48) = 7.09$, $p < .001$, $d = 1.01$; $BF_{10} = 2.28 \times 10^6$] than happy expressions [mean difference = .26, SD = .41; $t(48) = 4.46$, $p < .001$, $d = .64$; $BF_{10} = 435$]. As predicted, and as evidenced by a significant two-way interaction between Own-Pain and Precision [$F(1, 48) = 7.61$, $p = .008$, $\eta_p^2 = .14$], the effect of Own-Pain on ratings of pained expressions was greater in the high precision [mean difference = .50, *SD* = .50; $t(48) = 7.05$, $p < .001$, $d = 1.01$; $BF_{10} = 2.01 \times 10^6$] than the low precision [mean difference = .27, *SD* = .47; $t(48) = 4.07$, $p < .001$, $d = .58$; $BF_{10} = 139$] condition. Conversely, the simple main effect of Own-Pain on ratings of happy expressions was greater in the low precision [mean difference = .36, *SD* = .51; $t(48) = 4.90$, $p < .001$, $d = .70$; $BF_{10} = 1{,}689$] than the high precision [mean difference = .16, *SD* = .45; $t(48) = 2.49$, $p = .016$, $d = .36$; $BF_{10} = 2.48$] condition (see Fig. 3), and this interaction between Own-Pain and Precision was significant [$F(1, 48) = 7.10$, $p = .010$, $\eta_p^2 = .13$].

These results are confirmed by a one-tailed Bayesian paired samples *t*-test comparing the 2-way interaction effects (computed as the difference in the effect of pain on expression ratings between high and low precision conditions) for happy and pained expressions. A BF_{10} of 41 constitutes strong evidence for the predicted interaction between Pain, Precision and Emotion.

3.3. *Confirmatory and control analyses*

If the effect on expression intensity ratings is as predicted by the predictive processing framework (and Bayesian perception accounts in general), one would expect an effect of cue precision on the variance of own-pain ratings. Precise interoceptive predictions as to the intensity of the upcoming painful stimulation would be combined with sensory evidence to form a precise posterior distribution, leading to lower variance in reported own-pain given the same sensory evidence (i.e., to stimulations of equal intensity). Conversely, imprecise priors would be combined with sensory evidence to form an imprecise posterior distribution, and higher variance in own-pain perception for stimulations of equal intensity (see Hoskin et al., 2019). As a confirmatory analysis therefore, the variance of own-pain ratings was calculated for stimulations at the '3' and '8' levels (the two stimulation intensities shared in the precise and imprecise distributions) for each participant. Variance was calculated after equalising trial numbers in the precise and imprecise conditions by randomly sampling from the precise condition. These intensities were analysed using a one-tailed paired samples *t*-test which revealed a significant difference in the variance of own-pain ratings, $t(49) = 2.00$, $p = .026$, $d = .29$; $BF_{10} = 1.88$, although note that the Bayes factor provided only anecdotal evidence in favour of the alternative hypothesis (likely due to low power as a consequence of reduced trial numbers).

A control analysis was conducted to ensure that the observed effects were due to the precision of exteroceptive predictions (and therefore the degree to which exteroceptive predictions biased perception), rather than being a product of either of two alternative mechanisms. The first alternative is that the precision of interoceptive predictions affected the mean magnitude of experienced pain, with the relationship between experienced pain and expression intensity judgements remaining constant. The second alternative is that the emotional expression may have affected the experienced pain magnitude, since the predictive processing framework predicts bidirectional biasing effects whereby not only can the experience of pain cause an expression to be perceived as more pained to reduce exteroceptive prediction errors, but the sight of a pained expression can cause pain to be experienced as more intense to reduce interoceptive prediction errors. In order to rule out these alternative explanations, the own-pain ratings were analysed using the same 2 (Own-Pain: high *vs* low) × 2 (Precision: high *vs* low) × 2 (Emotion: pain *vs* happiness) repeated measures ANOVA used to analyse the expression intensity ratings, and supplemented with Bayesian paired samples *t*-tests. For interaction effects, the *t*-tests were performed on the relevant difference scores (e.g., for the Emotion × Own-Pain interaction, the differences in own-pain ratings between high and low own-pain were computed for both pain and happiness) in the same manner as above for expression ratings.

The ANOVA revealed no significant main effect of Precision [$F(1, 48) = 3.70$, $p = .060$, $\eta_p^2 = .072$; $BF_{10} = .85$]. While the frequentist ANOVA revealed a main effect of Emotion on experienced pain [$F(1, 48) = 7.75$, $p = .008$, $\eta_p^2 = .14$; $BF_{10} = 4.76$] such that own-pain was rated significantly higher when viewing pained faces ($M = 5.14$, $SD = .52$) than when viewing happy faces ($M = 5.06$, $SD = .58$), neither of the relevant 2-way interactions [Precision × Emotion: $F(1, 48) = .014$, $p = .907$, $\eta_p^2 = .0003$, $BF_{10} = .16$; Emotion × Own-Pain: $F(1, 48) = .78$, $p = .381$, $\eta_p^2 = .016$, $BF_{10} = .23$], nor the crucial three-way interaction were significant [$F(1, 48) = .009$, $p = .923$, $\eta_p^2 = .0002$; $BF_{10} = .16$]. The pattern of significance therefore does not suggest that the effects of either the precision of interoceptive cues or emotional stimulus on experienced own-pain explain the effect of the interoceptive cues on expression intensity ratings. Even if one ignores the pattern of significance and Bayes factors, given that a difference in own-pain ratings of 5 points was necessary to produce a mean difference of .32 in expression intensity ratings, it is unlikely that mean differences in own-pain approximately 90 times smaller than that between precision conditions, and 60 times smaller than between emotion conditions, could account for effects on expression intensity ratings.

4. Discussion

This study sought to test the hypothesis that the extent to which perception of another's expression is biased towards one's own state should vary as a function of the precision of predictions regarding one's own state. Results supported the hypothesis; increased precision of predictions about one's own pain was associated with increasingly biased perception of another's pain. Furthermore, the effect of the precision of

pain predictions on ratings of the intensity of happy expressions was reversed, such that less precise predictions were associated with the greatest effect on expression intensity ratings. This suggests that more precise predictions concerning one's own state enhance biasing effects of that state on the perception of a congruent emotion, and attenuate effects on the perception of an incongruent emotion.

The reason for the main effect of own-pain on happiness is unclear, but it is possible that high arousal states in the self enhance perception of emotions with a similar degree of arousal. Empirical evidence for such an effect in the domain of emotional valence (rather than arousal) is provided by Antico, Cataldo, and Corradi-Dell'Acqua (2019), who showed that a pained state enhances perception not only of pain but also, to a lesser degree, disgust (a similarly negatively-valenced state).

Hypotheses as to the effect of precision were based on the description of hierarchical generative models under predictive processing theories applied to interoception (e.g., Pezzulo, 2014; Pezzulo et al., 2015; Seth, 2013; Seth & Friston, 2016), in particular those extended to the interpersonal domain (Demekas et al., 2020; Ondobaka et al., 2017; Quattrocki & Friston, 2014). These models generate multimodal predictions and therefore can link exteroceptive, proprioceptive and interoceptive states. This property, combined with a developmental environment in which states of the self reliably predict, and are predicted by, states of the other, allow predictions concerning the other to influence the self and vice versa (Bird & Viding, 2014; Heyes & Bird, 2007; Ondobaka et al., 2017; Quattrocki & Friston, 2014). Such models are therefore consistent with the idea that learning resolves the 'correspondence problem' (whereby information about the state of another is typically acquired through exteroceptive senses such as vision and audition, while states of the self are typically encoded in interoceptive or proprioceptive codes) inherent in interpersonal influence (Cook, Bird, Catmur, Press, & Heyes, 2014).

While the present study has focused on perceptual inference, many of the theories described suggest that prediction error can also be resolved via active inference, whereby action is taken to alter sensory evidence such that it better matches predictions (e.g., Ondobaka et al., 2017; Quattrocki & Friston, 2014; Seth & Friston, 2016). According to these theories, active inference may explain how the state of the other can cause a change in one's own state, as in the case of emotion contagion or empathy. For example, if the sight of another in pain causes the perceiver's internal model of pain to generate autonomic and motor predictions, the ensuing prediction errors that occur if the perceiver is not themselves in a state congruent with pain can be resolved by engaging the autonomic and motor responses associated with pain (e.g., an increase in heart rate and electrodermal activity, muscle flexion), thus instantiating a state of empathic pain in the perceiver. In principle, the ensuing prediction errors may also be resolved through perceptual inference – a precise prior of pain perception in the self will bias perception of one's own state towards increasing pain – however recent research suggests that the influence of priors on pain perception is greatly reduced when prediction errors are large (Hird, Charalambous, El-Deredy, Jones, & Talmi, 2019).

The ability of interoceptive predictions to bias exteroceptive perception, as shown here, is consistent with accounts which suggest that interoception biases attentional, sensory and behavioural responses to stimuli that are homeostatically relevant (e.g., Barrett & Simmons, 2015). As argued by Seth and Friston (2016), the active inference framework highlights the relevance of predictive models to the regulation (not just prediction) of causes of sensory evidence. Due to their influence on our own states, the states of others are homeostatically relevant, and thus a target for regulation by predictive models.

In explaining how interpersonal influence arises, one must also explain how such effects can be overcome – why it is not the case that we compulsively mirror the autonomic/emotional states of others or imitate their actions (echopraxia), and why pairs of individuals do not become locked into such positive feedback loops. Active inference models posit that in order to avoid emotional echopraxia when confronted with another's pain, one must reduce the precision of interoceptive predictions and the ensuing descending prediction errors that would otherwise engage autonomic reflexes to perform the interoceptive action (Ondobaka et al., 2017; Quattrocki & Friston, 2014; Seth & Friston, 2016). With respect to the effect observed here – where the state of the self influences perception of another's state – one would need to reduce the precision of exteroceptive predictions.

Given the role of precision in the perception and prediction of others' states and in moderating interpersonal effects, it has been suggested that atypical precision-weighting may result in atypical social responses, with Autism Spectrum Disorder most frequently cited as a condition where this may be the case (Brock, 2012; Coll, Whelan, Catmur, & Bird, 2020; Pellicano & Burr, 2012; Quattrocki & Friston, 2014; but see Brewer, Happé, Cook, & Bird, 2015). In the case of empathy for pain, it has been argued that individuals with autism may fail to attenuate the precision of interoceptive predictions, leading to increased autonomic echopraxia and reduced cognitive understanding of others' pain (Gu et al., 2015; although see Bird et al., 2010).

However, even typical precision-weighting may be associated with sociocognitive difficulties in the absence of appropriate generative models linking predictions concerning the states of self and others. Since these models are a product of experience, they not only depend on sufficient and appropriate caregiver–child interaction (i.e., where the caregiver mirrors the child's states), but are also subject to individual, familial, and cultural variance (Conway, Catmur, & Bird, 2019; Demekas et al., 2020; Happé & Frith, 1996; Jack, Caldara, & Schyns, 2012; Russell, 1991; Smith, Parr, & Friston, 2019), meaning that predictive models may be appropriate for some individuals, or groups, but not others. Social interaction and communication with members of groups characterised by similar generative models as the self may therefore be easier than with those with different generative models (Edey et al., 2017; Friston & Frith, 2015; Keating & Cook, 2020; Schuster et al., 2021; Seth & Friston, 2016).

When interpreting the current findings, it is important to consider some potential limitations of the study. As participants were not asked to provide ratings of the expressions in the absence of any own-pain, the effect of pain on perception

of the expressions cannot be determined (only the relative effects of high and low pain). These data would have been useful in interpreting the effect of painful stimulation on perception of happiness. In addition, expressions were displayed on a single identity only, and so further research will be needed to confirm the generalisability of these effects. Finally, the use of morphed stimuli and static expressions limit the ecological validity of the study.

5. Conclusions

This study sought to test the hypothesis that perception of another's expression will be biased towards one's own state to a greater extent when predictions regarding one's own state are more precise. Results supported the hypothesis; increased precision of predictions about one's own pain was associated with increasingly biased perception of another's pain. These results are in accordance with predictive processing theories, which hypothesise the existence of hierarchical multimodal models that generate predictions concerning states of the self and others.

Author contributions

L. Sevi: Conceptualization, Methodology, Formal analysis, Investigation, Visualization, Writing – Original Draft.

M. Stantic: Methodology, Writing – Reviewing and Editing.

J. Murphy: Methodology, Writing – Reviewing and Editing.

M.P. Coll: Methodology, Writing – Reviewing and Editing.

C. Catmur: Methodology, Supervision, Writing – Reviewing and Editing.

G. Bird: Conceptualization, Methodology, Supervision, Writing – Reviewing and Editing.

Data availability

All data and code that support the findings of this study, including those presented in Fig. 3, have been deposited in the "Open Science Framework" (https://osf.io/4p5ur/). Stimuli cannot be shared as the original photographs are copyright of the research group who created them (Simon et al., 2008). The TAS-20 also cannot be shared as it is copyright of its original creators (Bagby et al., 1994).

Open practices

The study in this article earned an Open Data – Protected Access badge for transparent practices. Materials and data for the study are available at https://osf.io/4p5ur/.

Declaration of competing interest

None.

Acknowledgements

G. Bird was supported by an ESRC Grant (ES/R007527/1) and the Baily Thomas Charitable Trust.

REFERENCES

Allen, M., Levy, A., Parr, T., & Friston, K. J. (2019). In the body's eye: The computational anatomy of interoceptive inference. *BioRxiv*, Article 603928. https://doi.org/10.1101/603928

Antico, L., Cataldo, E., & Corradi-Dell'Acqua, C. (2019). Does my pain affect your disgust? Cross-modal influence of first-hand aversive experiences in the appraisal of others' facial expressions. *European Journal of Pain, 23*(7), 1283–1296. https://doi.org/10.1002/ejp.1390

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale–I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*(1), 23–32. https://doi.org/10.1016/0022-3999(94)90005-1

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews. Neuroscience, 16*(7), 419–429. https://doi.org/10.1038/nrn3950

Bird, G., Silani, G., Brindley, R., White, S., Frith, U., & Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain, 133*(5), 1515–1525. https://doi.org/10.1093/brain/awq060

Bird, G., & Viding, E. (2014). The self to other model of empathy: Providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience and Biobehavioral Reviews, 47*, 520–532. https://doi.org/10.1016/j.neubiorev.2014.09.021

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brewer, R., Cook, R., & Bird, G. (2016). Alexithymia: A general deficit of interoception. *Royal Society Open Science, 3*(10), Article 150664. https://doi.org/10.1098/rsos.150664

Brewer, R., Happé, F., Cook, R., & Bird, G. (2015). Commentary on "Autism, oxytocin and interoception": Alexithymia, not Autism Spectrum Disorders, is the consequence of interoceptive failure. *Neuroscience and Biobehavioral Reviews, 56*, 348–353. https://doi.org/10.1016/j.neubiorev.2015.07.006

Brock, J. (2012). Alternative Bayesian accounts of autistic perception: Comment on Pellicano and Burr. *Trends in Cognitive Sciences, 16*(12), 573–574. https://doi.org/10.1016/j.tics.2012.10.005. author reply 574.

Chapon, A., Perchet, C., Garcia-Larrea, L., & Frot, M. (2019). Hyperalgesia when observing pain-related images is a genuine bias in perception and enhances autonomic responses. *Scientific Reports, 9*(1), Article 15266. https://doi.org/10.1038/s41598-019-51743-3

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190217013.001.0001

Coll, M.-P., Whelan, E., Catmur, C., & Bird, G. (2020). Autistic traits are associated with atypical precision-weighted integration of top-down and bottom-up neural signals. *Cognition, 199*, Article 104236. https://doi.org/10.1016/j.cognition.2020.104236

Conway, J. R., Catmur, C., & Bird, G. (2019). Understanding individual differences in theory of mind via representation of minds, not mental states. *Psychonomic Bulletin & Review, 26*(3), 798–812. https://doi.org/10.3758/s13423-018-1559-x

Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. (2014). Mirror neurons: From origin to function. *Behavioral and Brain Sciences, 37*(2), 177–192. https://doi.org/10.1017/S0140525X13000903

Demekas, D., Parr, T., & Friston, K. J. (2020). An investigation of the free energy principle for emotion recognition. *Frontiers in Computational Neuroscience, 14*, 30. https://doi.org/10.3389/fncom.2020.00030

Edey, R., Yon, D., Cook, J., Dumontheil, I., & Press, C. (2017). Our own action kinematics predict the perceived affective states of others. *Journal of Experimental Psychology. Human Perception and Performance, 43*(7), 1263–1268. https://doi.org/10.1037/xhp0000423

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews. Neuroscience, 11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K., & Frith, C. (2015). A Duet for one. *Consciousness and Cognition, 36*, 390–405. https://doi.org/10.1016/j.concog.2014.12.003

Gu, X., Eilam-Stock, T., Zhou, T., Anagnostou, E., Kolevzon, A., Soorya, L., et al. (2015). Autonomic and brain responses associated with empathy deficits in autism spectrum disorder. *Human Brain Mapping, 36*(9), 3323–3338. https://doi.org/10.1002/hbm.22840

Happé, F., & Frith, U. (1996). Theory of mind and social impairment in children with conduct disorder. *British Journal of Developmental Psychology, 14*(4), 385–398. https://doi.org/10.1111/j.2044-835X.1996.tb00713.x

Heyes, C., & Bird, G. (2007). Mirroring, association, and the correspondence problem. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Sensorimotor foundations of higher cognition* (pp. 460–480). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199231447.003.0021.

Hird, E. J., Charalambous, C., El-Deredy, W., Jones, A. K. P., & Talmi, D. (2019). Boundary effects of expectation in human pain perception. *Scientific Reports, 9*(1), 1–13. https://doi.org/10.1038/s41598-019-45811-x

Hohwy, J. (2013). *The predictive mind*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199682737.001.0001

Hoskin, R., Berzuini, C., Acosta-Kane, D., El-Deredy, W., Guo, H., & Talmi, D. (2019). Sensitivity to pain expectations: A Bayesian model of individual differences. *Cognition, 182*, 127–139. https://doi.org/10.1016/j.cognition.2018.08.022

Jack, R. E., Caldara, R., & Schyns, P. G. (2012). Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *Journal of Experimental Psychology: General, 141*(1), 19–25. https://doi.org/10.1037/a0023463

Jasp Team. (2020). *JASP* (Version 0.14.1) [Computer software] https://jasp-stats.org/.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

Keating, C. T., & Cook, J. L. (2020). Facial expression production and recognition in autism spectrum disorders. *Child and Adolescent Psychiatric Clinics of North America, 29*(3), 557–571. https://doi.org/10.1016/j.chc.2020.02.006

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing, 8*(3), 159–166. https://doi.org/10.1007/s10339-007-0170-2

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: A neural prediction problem. *Neuron, 79*(5), 836–848. https://doi.org/10.1016/j.neuron.2013.08.020

Linson, A., Parr, T., & Friston, K. J. (2020). Active inference, stressors, and psychological trauma: A neuroethological model of (mal) adaptive explore-exploit dynamics in ecological context. *Behavioural Brain Research, 380*, Article 112421. https://doi.org/10.1016/j.bbr.2019.112421

Liu, Y., Meng, J., Yao, M., Ye, Q., Fan, B., & Peng, W. (2019). Hearing other's pain is associated with sensitivity to physical pain: An ERP study. *Biological Psychology, 145*, 150–158. https://doi.org/10.1016/j.biopsycho.2019.03.011

Mirza, M. B., Cullen, M., Parr, T., Shergill, S., & Moran, R. J. (2021). Contextual perception under active inference. *Scientific Reports, 11*(1), 1–25. https://doi.org/10.1038/s41598-021-95510-9

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *The Quantitative Methods for Psychology, 4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Morpheus Development. Morpheus Photo Morpher (Version 3.17) [Computer software] http://www.morpheussoftware.net/.

Ondobaka, S., Kilner, J., & Friston, K. (2017). The role of interoceptive inference in theory of mind. *Brain and Cognition, 112*, 64–68. https://doi.org/10.1016/j.bandc.2015.08.002

Pellicano, E., & Burr, D. (2012). When the world becomes "too real": A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences, 16*(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

Peng, W., Huang, X., Liu, Y., & Cui, F. (2019). Predictability modulates the anticipation and perception of pain in both self and others. *Social Cognitive and Affective Neuroscience, 14*(7), 747–757. https://doi.org/10.1093/scan/nsz047

Pezzulo, G. (2014). Why do you fear the bogeyman? An embodied predictive coding model of perceptual inference. *Cognitive, Affective & Behavioral Neuroscience, 14*(3), 902–911. https://doi.org/10.3758/s13415-013-0227-x

Pezzulo, G., Iodice, P., Barca, L., Chausse, P., Monceau, S., & Mermillod, M. (2018). Increased heart rate after exercise facilitates the processing of fearful but not disgusted faces. *Scientific Reports, 8*(1), 398. https://doi.org/10.1038/s41598-017-18761-5

Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology, 134*, 17–35. https://doi.org/10.1016/j.pneurobio.2015.09.001

Quattrocki, E., & Friston, K. (2014). Autism, oxytocin and interoception. *Neuroscience and Biobehavioral Reviews, 47*, 410–430. https://doi.org/10.1016/j.neubiorev.2014.09.012

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*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological Bulletin, 110*(3), 426–450. https://doi.org/10.1037/0033-2909.110.3.426

Rütgen, M., Seidel, E.-M., Silani, G., Riečanský, I., Hummer, A., Windischberger, C., et al. (2015). Placebo analgesia and its opioidergic regulation suggest that empathy for pain is grounded in self pain. *Proceedings of the National Academy of Sciences of the United States of America, 112*(41), E5638–E5646. https://doi.org/10.1073/pnas.1511269112

Rütgen, M., Wirth, E.-M., Riečanský, I., Hummer, A., Windischberger, C., Petrovic, P., et al. (2021). Beyond sharing unpleasant affect-evidence for pain-specific opioidergic modulation of empathy for pain. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhaa385

Schuster, B. A., Fraser, D. S., van den Bosch, J. J. F., Sowden, S., Gordon, A. S., Huh, D., et al. (2021). Kinematics and observer-animator kinematic similarity predict mental state attribution from Heider–Simmel style animations. *Scientific Reports, 11*(1), 1–16. https://doi.org/10.1038/s41598-021-97660-2

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences, 17*(11), 565–573. https://doi.org/10.1016/j.tics.2013.09.007

Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 371*(1708). https://doi.org/10.1098/rstb.2016.0007

Silani, G., Lamm, C., Ruff, C. C., & Singer, T. (2013). Right supramarginal gyrus is crucial to overcome emotional egocentricity bias in social judgments. *The Journal of Neuroscience, 33*(39), 15466–15476. https://doi.org/10.1523/JNEUROSCI.1488-13.2013

Simon, D., Craig, K. D., Gosselin, F., Belin, P., & Rainville, P. (2008). Recognition and discrimination of prototypical dynamic expressions of pain and emotions. *Pain, 135*(1–2), 55–64. https://doi.org/10.1016/j.pain.2007.05.008

Smith, R., Parr, T., & Friston, K. J. (2019). Simulating emotions: An active inference model of emotional state inference and emotion concept learning. *Frontiers in Psychology, 10*, 2844. https://doi.org/10.3389/fpsyg.2019.02844