

Review of the PhD Thesis of Sylvia Boschetti

RNDr James G. Pfaus, PhD

Department of Psychology and Life Sciences, Faculty of Humanities, Charles University
Center for Sexual Health and Intervention, Czech National Institute of Mental Health

This thesis reviews several papers published by the candidate and her collaborators and co-workers on the ability of women and men to show accurate visual pattern recognition, and the ability of auditory stimuli to aid or detract from that. In the first sets of studies, the candidate utilized a commercially-available automaton to create sets of visual stimuli in color or in black-and-white to study false positive errors in pattern detection (pareidolia) and false negative errors (apoidolia). The automaton created sets of stimuli along a gradient from random to non-random, and of varying difficulty for the non-random stimuli using random masking and uniformly colored shapes. This eliminated the need for a pre-test of inter-rater reliability in pattern detection. Participants answered questionnaires concerning their thinking styles (e.g., rational, intuitive). This set of studies found differences in participants that accurately rated randomness and non-randomness, and those who showed a high degree of false positives versus false negatives, differences that were related in a cluster analysis to thinking style.

A second set of the candidate's studies was reviewed that extended this analysis to the ability of women and men to properly process the affective states of pain, pleasure, fear, laugh, and a neutral control, based on visual and/or auditory signaling. Visual stimuli restricted to the face, and concurrent auditory stimuli, were selected from a large number of visual sexual stimuli that represented the 5 affective states, and then equated for luminosity, etc. To reduce noise, stimuli were further evaluated and finally selected for use on the basis of an AI-generated "neural net" autoencoder. Appropriate Bayesian statistical methods were utilized for analyses. The AI algorithm was able to distinguish the five affective states based on a single sensory modality better than humans, and this again eliminated the need for a pre-test of inter-rater reliability. Participants could correctly identify laugh and neutral, less so for fear, and were at chance for pain and pleasure, based on facial features with or without the corresponding auditory component. In contrast, participants were less likely to guess from auditory stimuli alone (relative to visual stimuli) but the two together produced greater confidence in the participants' decision in the congruent condition than when rating facial expressions alone, but lower confidence than when rating vocalizations alone. And when the visual and auditory stimuli were made incongruent, most participants based their ratings on the auditory condition,

suggesting that it suppressed the visual. Finally, the effect of stress and arousal on these responses was examined. Participants were administered either a cold-pressor task (immersion of a lower hind limb in 2-4°C ice water for 90 sec) or a control condition in which the water was room temperature. This task is known to increase both blood pressure and cortisol responses in participants, although neither were assessed in the study. The task produced a significant increase in accuracy for the participants' rating of laughter and pleasure, but only for images of males. The opposite occurred for vocalizations. The task significantly decreased the probability of correct attribution in the stressed group for laugh and neutral, but also only in the case of images of males.

The candidate concludes her thesis with a reiteration of what was found, the potential of her methodology for AI facial feature detection (and ostensibly detection or prediction of a person's emotional state), and suggestions that her methodology avoids arguments of ecological validity by avoiding the bias inherent in studies that rely on pre-tested databases, FACS models, Likert scales, etc.

Certainly, this is an enormous amount of work, and the candidate has crafted an interesting and well-written thesis around it. Her work makes a clear contribution and I would recommend that the thesis move forward to a defense, the date of which has already been set for 18 April 2023.

There are some issues that the thesis should revise before submission, and that the candidate should address at her defense.

For example, the introduction goes into much detail about the evolutionary significance of feature detection and its errors (i.e., better to think the stick is a snake and run away than to think the snake is a stick and ignore it), which is entirely appropriate. But it was curious to me that a thorough review of the rather large literature on human detection of emotion in faces was not provided, nor was there any attempt to place the findings into that more general literature in the conclusions. For example, humans are better at assessing facial features that denote anger (threat) relative to happy, and this kind of Gestalt feature detection serves the immediate protection of the individual. And it is very true that facial features during orgasm resemble those of pain, more so than facial features of other emotional states, including sexual arousal and desire. Now, I understand that the thesis was more about developing a relatively unbiased methodology and the use of sophisticated Bayesian statistical techniques to assess feature detection errors, so I wonder if this was an oversight and to what extent the candidate knows that literature and the implications of her findings within it.

Like other animals, humans must make split-second decisions based on information from multiple sensory modalities, and indeed, within contexts. All of these are Gestalts, to some extent, that allow meaning to "pop out" from the background noise of all sensory information

being encoded. And although it is true that humans rely mostly on our sense of vision and audition, it is also true that ecological validity comes with more, not less, sensory information, even in circumstances that are new, unconditioned, and/or unpredictable. There is more to emotional detection than facial muscle movements and vocalizations captured in a static second. Context plays an important role. One of the exciting possibilities of the methodology used here is in the realm of a set of “add back” experiments (like old Wundtian reaction time experiments), in which more and more features (and potentially for greater and greater lengths of time) are added to the faces and vocalizations until the probability of a correct assessment increases to 100%. Could the candidate envision such a set of experiments and what might the critical feature(s) that correct assessment actually depends on?

From a neurobiological perspective, why might auditory signals be more linked to emotion than visual signals, even those visual signals that depict emotional facial features?

The Cold Pressor Task was used to generate a state of arousal. What do you think other arousal states (especially states of eustress) might do to the processing accuracy of the same visual and/or auditory signals (both abstract and emotional) used in the present thesis? Finally, why does the candidate think this task only had effects on the emotional processing of male faces and vocalizations?

The first chapter of the thesis (Overview) contains no references. I would imagine that something like Error Management Theory, Lady Tasting Tea, and the Cold Pressor Task (which is erroneously referred to as the “Cold Pressure Task” (P. 13) should have references, along with the candidate’s own publications that constitute the body of work the thesis reviews. Likewise, the second chapter cites many phenomena without references. It is not until the candidate cites Uexkull (Pp. 18-19) that proper referencing starts to appear.

On P. 21, the candidate refers to modality cross talk. Is this a sense organ issue or a brain neural interpretation problem? And at what level of processing (e.g., primary input to sense organ, secondary (thalamic) input, tertiary input to the cortex, and/or subsequent integrative processing in association cortex)?

P. 22: How do cues enable animals to navigate environments? How do animals respond to cues? Are signals pre-potent and cues learned, or are signals learned as well (especially when coming from different sensory modalities)? This makes me wonder if something like Discrete Category Theory and the Circumplex Model of Emotions could actually act together, albeit differently in different brain regions (e.g., limbic system versus cortex).

On P. 28, Figure 3-1 is labelled Figure 2-1. It is stated that the incoming data are processed sequentially by primary, secondary, and tertiary AI analytics, which ultimately reduce the data

to the activation of two orange “neurons.” However, the figure depicts three orange “neurons.” Why is this?

P. 30: Are guesses not based on available information, including past experiences and the willingness to take risks?

P. 40: Is “laugh” an affective state, or a response to the induction of an affective state? If the latter, would the affective state be amusement, humor, or happiness (or any combination)? And regarding “laugh”, why was it considered a low intensity affective state relative to pleasure or pain (especially given its ability to be infectious in groups of people)?

On P. 46, it is stated that arousal is a function of the sympathetic nervous system. Is that true of all arousal states? What about sexual arousal?

Why were the “fear” results in the candidate’s papers not discussed in the thesis?

Why does the candidate think the AI algorithm is better at detecting emotion in facial features and vocalizations than the human brain?

P. 50, second-to-last paragraph: How has the thesis connected evolutionary biological, psychological, and ethological approaches? Please elaborate.

Signed in Praha on 12 April 2023/12. dubna 2023:



RNDr James G. Pfaus, PhD